PATENT APPLICATION
Attorney Docket No. 15966-556 CIP1 (Cura 56 CIP1)

POLYPEPTIDES AND POLYNUCLEOTIDES ENCODING SAME

RELATED APPLICATIONS

This application is a continuation-in-part of USSN 09/619252 filed July 19, 2000, which claims priority to USSN 60/144,722, filed July 20, 1999, and USSN 60/167,785, filed November 29, 1999; and is a continuation-in-part of USSN 60/276,994 filed March 19, 2001; USSN 60/280898 filed April 2, 2001; USSN 60/332,241 filed November 14, 2001; USSN 60/288,062 filed May 2, 2001; USSN 60/291,766 filed May 17, 2001; and USSN 60/314,007 filed August 21, 2001. The contents of these applications are incorporated herein by reference in their entireties.

FIELD OF THE INVENTION

The invention relates to generally to polynucleotides and the polypeptides encoded thereby and more particularly to polynucleotides encoding polypeptides that cross one or more membranes in eukaryotic cells.

BACKGROUND OF THE INVENTION

Eukaryotic cells are subdivided by membranes into multiple, functionally-distinct compartments, referred to as organelles. Many biologically important proteins are secreted from the cell after crossing multiple membrane-bound organelles. These proteins can often be identified by the presence of sequence motifs referred to as "sorting signals" in the protein, or in a precursor form of the protein. These sorting signals can also aid in targeting the proteins to their appropriate destination.

One specific type of sorting signal is a signal sequence, which is also referred to as a signal peptide or leader sequence. This signal sequence, which can be present as an amino-terminal extension on a newly synthesized polypeptide. A signal sequence possesses the ability to "target" proteins to an organelle known as the endoplasmic reticulum (ER).

The signal sequence takes part in an array of protein-protein and protein-lipid interactions that result in the translocation of a signal sequence-containing polypeptide through a channel within the ER. Following translocation, a membrane-bound enzyme, designated signal peptidase, liberates the mature protein from the signal sequence.

Secreted and membrane-bound proteins are involved in many biologically diverse activities. Examples of known, secreted proteins include, *e.g.*, insulin, interferon, interleukin,

transforming growth factor-β, human growth hormone, erythropoietin, and lymphokine. Only a limited number of genes encoding human membrane-bound and secreted proteins have been identified.

Failure to thrive, nutritional edema, and hypoproteinemia with normal sweat electrolytes of 2 affected male infants reported by Townes et al (J. Pediat. 71: 220-224, 1967), could be treated by a protein hydrolysate diet. Morris and Fisher (Am. J. Dis. Child. 114: 203-208, 1967) reported an affected female who also had imperforate anus, a result of a defect in the synthesis of the enterokinase which activates proteolytic enzymes produced by the pancreas. Oral pancreatin represents a therapeutically successful form of enzyme replacement. Trypsin, like elastase is a member of the pancreatic family of serine proteases. MacDonald et al. (J. Biol. Chem. 257: 9724-9732, 1982) reported nucleotide sequences of cDNAs representing 2 pancreatic rat trypsinogens. The trypsin gene is on mouse chromosome 6 (Honey et al., Somat. Cell Molec. Genet. 10: 369-376, 1984). Carboxypeptidase A and trypsin are a syntenic pair conserved in mouse and man. Emi et al. (Gene 41: 305-310, 1986) isolated cDNA clones for 2 major human trypsinogen isozymes from a pancreatic cDNA library. The deduced amino acid sequences had 89% homology and the same number of amino acids (247), including a 15-amino acid signal peptide and an 8-amino acid activation peptide. Southern blot analysis of human genomic DNA with the cloned cDNA as a probe showed that the human trypsinogen genes constitute a family of more than 10. The gene encoding trypsin-1 (TRY1) is also referred to as serine protease-1 (PRSS1). Rowen et al. (Science 272: 1755-1762, 1996) found that there are 8 trypsinogen genes embedded in the beta T-cell receptor locus or cluster of genes (TCRB) mapping to 7q35. In the 685-kb DNA segment that they sequenced they found 5 tandemly arrayed 10-kb locus-specific repeats (homology units) at the 3-prime end of the locus. These repeats exhibited 90 to 91% overall nucleotide similarity, and embedded within each is a trypsinogen gene. Alignment of pancreatic trypsinogen cDNAs with the germline sequences showed that these trypsinogen genes contain 5 exons that span approximately 3.6 kb. They denoted 8 trypsinogen genes T1 through T8 from 5-prime to 3-prime. Some of the trypsinogen genes are expressed in nonpancreatic tissues where their function is unknown. Rowen et al. (Science 272: 1755-1762, 1996) noted that the intercalation of the trypsinogen genes in the TCRB locus is conserved in mouse and chicken, suggesting shared functional or regulatory constraints, as has been postulated for genes in the major histocompatibility complex (such as class I, II, and III genes) that share similar long-term

organizational relationships. The gene of invention is a novel serine protease containing a trypsin domain but localized on chromosome 16.

SUMMARY OF THE INVENTION

The invention is based, in part, upon the discovery of novel nucleic acids and secreted polypeptides encoded thereby. The nucleic acids and polypeptides are collectively referred to herein as "SECP".

Accordingly, in one aspect, the invention includes an isolated nucleic acid that encodes a SECP polypeptide, or a fragment, homolog, analog or derivative thereof. For example, the nucleic acid can encode a polypeptide at least 85% identical to a polypeptide comprising the amino acid sequences of SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 41, 43, 45, 47, 49, 51, 53, 55 and 57. The nucleic acid can be, *e.g.,* a genomic DNA fragment, cDNA molecule. In some embodiments, the nucleic acid includes the sequence the invention provides an isolated nucleic acid molecule that includes the nucleic acid sequence of any of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and 56.

Also included within the scope of the invention is a vector containing one or more of the nucleic acids described herein, and a cell containing the vectors or nucleic acids described herein.

The invention is also directed to host cells transformed with a vector comprising any of the nucleic acid molecules described above.

In another aspect, the invention includes a pharmaceutical composition that includes a SECP nucleic acid and a pharmaceutically acceptable carrier or diluent.

In a further aspect, the invention includes a substantially purified SECP polypeptide, *e.g.,* any of the SECP polypeptides encoded by a SECP nucleic acid, and fragments, homologs, analogs, and derivatives thereof. The invention also includes a pharmaceutical composition that includes a SECP polypeptide and a pharmaceutically acceptable carrier or diluent.

In a still a further aspect, the invention provides an antibody that binds specifically to a SECP polypeptide. The antibody can be, *e.g.,* a monoclonal or polyclonal antibody, and fragments, homologs, analogs, and derivatives thereof. The invention also includes a pharmaceutical composition including SECP antibody and a pharmaceutically acceptable carrier or diluent. The invention is also directed to isolated antibodies that bind to an epitope on a polypeptide encoded by any of the nucleic acid molecules described above.

The invention also includes kits comprising any of the pharmaceutical compositions described above.

The invention further provides a method for producing a SECP polypeptide by providing a cell containing a SECP nucleic acid, *e.g.*, a vector that includes a SECP nucleic acid, and culturing the cell under conditions sufficient to express the SECP polypeptide encoded by the nucleic acid. The expressed SECP polypeptide is then recovered from the cell. Preferably, the cell produces little or no endogenous SECP polypeptide. The cell can be, *e.g.*, a prokaryotic cell or eukaryotic cell.

The invention is also directed to methods of identifying a SECP polypeptide or nucleic acids in a sample by contacting the sample with a compound that specifically binds to the polypeptide or nucleic acid, and detecting complex formation, if present.

The invention further provides methods of identifying a compound that modulates the activity of a SECP polypeptide by contacting SECP polypeptide with a compound and determining whether the SECP polypeptide activity is modified.

The invention is also directed to compounds that modulate SECP polypeptide activity identified by contacting a SECP polypeptide with the compound and determining whether the compound modifies activity of the SECP polypeptide, binds to the SECP polypeptide, or binds to a nucleic acid molecule encoding a SECP polypeptide.

In a another aspect, the invention provides a method of determining the presence of or predisposition of a SECP-associated disorder in a subject. The method includes providing a sample from the subject and measuring the amount of SECP polypeptide in the subject sample. The amount of SECP polypeptide in the subject sample is then compared to the amount of SECP polypeptide in a control sample. An alteration in the amount of SECP polypeptide in the subject protein sample relative to the amount of SECP polypeptide in the control protein sample indicates the subject has a tissue proliferation-associated condition. A control sample is preferably taken from a matched individual, *i.e.*, an individual of similar age, sex, or other general condition but who is not suspected of having a tissue proliferation-associated condition. Alternatively, the control sample may be taken from the subject at a time when the subject is not suspected of having a tissue proliferation-associated disorder. In some embodiments, the SECP is detected using a SECP antibody.

In a further aspect, the invention provides a method of determining the presence of or predisposition of a SECP-associated disorder in a subject. The method includes providing a

nucleic acid sample (*e.g.*, RNA or DNA, or both) from the subject and measuring the amount of the SECP nucleic acid in the subject nucleic acid sample. The amount of SECP nucleic acid sample in the subject nucleic acid is then compared to the amount of a SECP nucleic acid in a control sample. An alteration in the amount of SECP nucleic acid in the sample relative to the amount of SECP in the control sample indicates the subject has a tissue proliferation-associated disorder.

In a still further aspect, the invention provides method of treating or preventing or delaying a SECP-associated disorder. The method includes administering to a subject in which such treatment or prevention or delay is desired a SECP nucleic acid, a SECP polypeptide, or a SECP antibody in an amount sufficient to treat, prevent, or delay a tissue proliferation-associated disorder in the subject.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In the case of conflict, the present Specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

BRIEF DESCRIPTION OF THE FIGURES

FIG. 1 is a representation of a SECP 1 nucleic acid sequence (SEQ ID NO:1) according to the invention, along with an amino acid sequence (SEQ ID NO:2) encoded by the nucleic acid sequence.

FIG. 2 is a representation of a SECP 2 nucleic acid sequence (SEQ ID NO:3) according to the invention, along with an amino acid sequence (SEQ ID NO:4) encoded by the nucleic acid sequence.

FIG. 3 is a representation of a SECP 3 nucleic acid sequence (SEQ ID NO:5) according to the invention, along with an amino acid sequence (SEQ ID NO:6) encoded by the nucleic acid sequence.

FIG. 4 is a representation of a SECP 4 nucleic acid sequence (SEQ ID NO:7) according to the invention, along with an amino acid sequence (SEQ ID NO:8) encoded by the nucleic acid sequence.

FIG. 5 is a representation of a SECP 5 nucleic acid sequence (SEQ ID NO:9) according to the invention, along with an amino acid sequence (SEQ ID NO:10) encoded by the nucleic acid sequence.

FIG. 6 is a representation of a SECP 6 nucleic acid sequence (SEQ ID NO:11) according to the invention, along with an amino acid sequence (SEQ ID NO:12) encoded by the nucleic acid sequence.

FIG. 7 is a representation of a SECP 7 nucleic acid sequence (SEQ ID NO:13) according to the invention, along with an amino acid sequence (SEQ ID NO:14) encoded by the nucleic acid sequence.

FIG. 8 is a representation of a SECP 8 nucleic acid sequence (SEQ ID NO:15) according to the invention, along with an amino acid sequence (SEQ ID NO:16) encoded by the nucleic acid sequence.

FIG. 9 is a representation of a SECP 9 nucleic acid sequence (SEQ ID NO:17) according to the invention, along with an amino acid sequence (SEQ ID NO:18) encoded by the nucleic acid sequence.

FIG. 10 is a representation of an alignment of the proteins encoded by clones 11618130.0.27 (SEQ ID NO:4) and 11618130.0.184 (SEQ ID NO:16).

FIG. 11 is a representation of an alignment of the proteins encoded by clones 14578444.0.143 (SECP4; SEQ ID NO:8) and 14578444.0.47 (SECP 5; SEQ ID NO:10).

FIG. 12 is a representation of a Western blot of a polypeptide expressed in 293 cells of a polynucleotide containing sequences encoded by clone 11618130.

FIG. 13 is a representation of a Western blot of a polypeptide expressed in 293 cells of a polynucleotide containing sequence encoded by clone 16406477.

FIG. 14 is a representation of a real-time expression analysis of the clones of the invention.

FIG. 15 is a representation of a SECP 10 nucleic acid sequence (SEQ ID NO:40) according to the invention, along with an amino acid sequence (SEQ ID NO:41) encoded by the nucleic acid sequence.

FIG. 16 is a representation of a SECP 11 nucleic acid sequence (SEQ ID NO:42) according to the invention, along with an amino acid sequence (SEQ ID NO:43) encoded by the nucleic acid sequence.

FIG. 17 is a representation of a SECP 12 nucleic acid sequence (SEQ ID NO:44) according to the invention, along with an amino acid sequence (SEQ ID NO:45) encoded by the nucleic acid sequence.

FIG. 18 is a representation of a SECP 13 nucleic acid sequence (SEQ ID NO:46) according to the invention, along with an amino acid sequence (SEQ ID NO:47) encoded by the nucleic acid sequence.

FIG. 19 is a representation of a SECP 14 nucleic acid sequence (SEQ ID NO:48) according to the invention, along with an amino acid sequence (SEQ ID NO:49) encoded by the nucleic acid sequence.

FIG. 20 is a representation of a SECP 15 nucleic acid sequence (SEQ ID NO:50) according to the invention, along with an amino acid sequence (SEQ ID NO:51) encoded by the nucleic acid sequence.

FIG. 21 is a representation of a SECP 16 nucleic acid sequence (SEQ ID NO:52) according to the invention, along with an amino acid sequence (SEQ ID NO:53) encoded by the nucleic acid sequence.

FIG. 22 is a representation of a SECP 17 nucleic acid sequence (SEQ ID NO:54) according to the invention, along with an amino acid sequence (SEQ ID NO:55) encoded by the nucleic acid sequence.

FIG. 23 is a representation of a SECP 18 nucleic acid sequence (SEQ ID NO:56) according to the invention, along with an amino acid sequence (SEQ ID NO:57) encoded by the nucleic acid sequence.

DETAILED DESCRIPTION OF THE INVENTION

The invention provides novel polynucleotides and the polypeptides encoded thereby. Included in the invention are ten novel nucleic acid sequences and their encoded polypeptides. These sequences are collectively referred to as "SECP nucleic acids" or "SECP polynucleotides" and the corresponding encoded polypeptide is referred to as a "SECP polypeptide" or "SECP protein". For example, a SECP nucleic acid according to the invention is a nucleic acid including a SECP nucleic acid, and a SECP polypeptide according to the invention is a polypeptide that includes the amino acid sequence of a SECP polypeptide. Unless indicated otherwise, "SECP" is meant to refer to any of the novel sequences disclosed herein. Each of the nucleic acid and amino acid sequences have been assigned a unique SECP Identification Number, with designations SECP1 through SECP10.

TABLE 1 provides a cross-reference to the assigned SECP Number, Clone or Probe Identification Number, and Sequence Identification Number (SEQ ID NO:) for both the nucleic acid and encoded polypeptides of SECP1-14.

TABLE 1

CLONE/PROBE	FIGURE	SEQ ID NO: (Nucleic Acid)	SEQ ID NO: (Polypeptide)
21433858	1	1	2
11618130.0.27, also called CG50817-03	2	3	4
11696905-0-47	3	5	6
14578444.0.143	4	7	8
14578444.0.47	5	9	10
14998905.0.65	6	11	12
16406477.0.206	7	13	14
11618130.0.184	8	15	16
21637262.0.64	9	17	18
CG106318-01	15	40	41
CG50817-04	16	42	43
CG50817-05	17	44	45
CG50817-06	18	46	47
CG51099-03	19	48	49
CG57051-04	20	50	51
CG57051-05	21	52	53
CG57051-02	22	54	55
CG57051-03	23	56	57
11618130 Forward		19	
11618130 Reverse		20	
PSec-V5-His Forward		21	
PSec-V5-His Reverse		22	
16406477 Forward		23	
16406477 Reverse		24	

Ag 383 (F)		25	
Ag 383 (R)		26	
Ag 383 (P)		27	
Ag 53 (F)		28	
Ag 53 (R)		29	
Ag 53 (P)		30	
Ag 127 (F)		31	
Ag 127 (R)		32	
Ag 127 (P)		33	
Ab 5(F)		34	
Ab 5(R)		35	
Ab 5(P)		36	
Ag 815(F)		37	
Ag 815(R)		38	
Ag 815(P)		39	

Nucleic acid sequences and polypeptide sequences for SECP nucleic acids and polypeptides, as disclosed herein, are provided in the following section of the Specification.

SECP nucleic acids, and their encoded polypeptides, according to the invention are useful in a variety of applications and contexts. For example, various SECP nucleic acids and polypeptides according to the invention are useful, *inter alia*, as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins.

SECP nucleic acids and polypeptides according to the invention can also be used to identify cell types based on the presence or absence of various SECP nucleic acids according to the invention. Additional utilities for SECP nucleic acids and polypeptides are discussed below.

SECP1

A SECP1 nucleic acid and polypeptide according to the invention includes the nucleic acid sequence (SEQ ID NO:1) and encoded polypeptide sequence (SEQ ID NO:2) of clone 21433858. FIG. 1 illustrates the nucleic acid and amino acid sequences, as well as the alignment between these two sequences.

This clone includes a nucleotide sequence (SEQ ID NO:1) of 6373 bp. The nucleotide sequence includes an open reading frame (ORF) encoding a polypeptide of 1588 amino acid residues (SEQ ID NO:2) with a predicted molecular weight of 178042.1 Daltons. The start codon is located at nucleotides 235-237 and the stop codon is located at nucleotides 4999-5001. The protein encoded by clone 21433858 is predicted by the PSORT program to localize in the plasma membrane with a certainty of 0.7300. The program SignalP predicts that there is a signal

peptide with the most probable cleavage site located between residues 23 and 24, in the sequence CMG-DE.

Real-time gene expression analysis was performed on SECP1 (clone 21433858). The results demonstrate that RNA sequences with homology to clone 21433858 are detected in various cell types. The relative abundance of RNA homologous to clone 21433858 is shown in FIG. 14 (see also Examples, below). Cell types endothelial cells (treated and untreated), pancreas, adipose, adrenal gland, thyroid, mammary gland, myometrium, uterus, placenta, prostate, testis, and in neoplastic cells derived from ovarian carcinoma OVCAR-3, ovarian carcinoma OVCAR-5, ovarian carcinoma OVCAR-8, ovarian carcinoma IGROV-1, ovarian carcinoma (ascites) SK-OV-3, breast carcinoma BT-549, prostate carcinoma (bone metastases) PC-3, Melanoma M14, and melanoma (met) SK-MEL-5. Accordingly, SECP1 nucleic acids according to the invention can be used to identify one or more of these cell types. The presence of RNA sequences homologous to a SECP1 nucleic in a sample indicates that the sample contains one or more of the above-cell types.

A search of sequence databases using BLASTX reveals that residues 299-1588 of the polypeptide encoded clone 21433858 are 100% identical to the 1290 residue human KIAA0960 protein (ACC: SPTREMBL-ACC:Q9UPZ6). In addition, the protein of clone 21433858 has 542 of 543 residues (99%) identical to, and 543 of 543 residues (100%) positive with, the 543 residue fragment of a human hypothetical protein (SPTREMBL-ACC:O60407).

The proteins of the invention encoded by clone 21433858 include the protein disclosed as being encoded by the ORF described herein, as well as any mature protein arising therefrom as a result of post-translational modifications. Thus, the proteins of the invention encompass both a precursor and any active forms of the clone 21433858 protein.

SECP2

A SECP2 nucleic acid and polypeptide according to the invention includes a nucleic acid sequence (SEQ ID NO:3) and an encoded polypeptide sequence (SEQ ID NO:4) of clone 11618130.0.27, also called CG50817-03. FIG. 2 illustrates the nucleic acid sequence and amino acid sequence, as well as the alignment between these two sequences.

This clone includes a nucleotide sequence (SEQ ID NO:3) of 1894 nucleotides. The nucleotide sequence includes an open reading frame (ORF) encoding a polypeptide of 267 amino acid residues with a predicted molecular weight of 28043 Daltons. The start codon is at nucleotides 732-734 and the stop codon is at nucleotides 1534-1536. The protein encoded by

clone 11618130.0.27 is predicted by the PSORT program to localize in the microbody (peroxisome) with a certainty of 0.5035. The program SignalP predicts that there is no signal peptide in the encoded polypeptide.

A search of the sequence databases using BLAST P and BLASTX reveals that clone 11618130.0.27 has 330 of 333 residues (99%) identical to and positive with a 571 residue human protein termed PRO351 (PCT Publication WO9946281-A2 published September 16, 1999). In addition, it was found to have 83 of 250 residues (33%) identical to, and 119 of 250 residues (47%) positive with the 343 residue human prostasin precursor (EC 3.4.21.-) (SWISSPROT-ACC:Q16651).

The proteins of the invention encoded by clone 11618130.0.27 includes the protein disclosed as being encoded by the ORF described herein, as well as any mature protein arising therefrom as a result of post-translational modification. Thus, the protein of the invention encompasses both a precursor and any active forms of the 11618130.0.27 protein.

SECP3

A SECP3 nucleic acid and polypeptide according to the invention includes the nucleic acid sequence (SEQ ID NO:5) and encoded polypeptide sequence (SEQ ID NO:6) of clone 11696905-0-47. FIG. 3 illustrates the nucleic acid sequence and amino acid sequence, as well as the alignment between these two sequences.

Clone 11696905-0-47 was obtained from fetal brain. In addition, RNA sequences were also found to be present in tissues including, uterus, pregnant and non-pregnant uterus, ovarian tumor, placenta, bone marrow, hippocampus, synovial membrane, fetal heart, fetal lung, pineal gland and melanocytes. This clone includes a nucleotide sequence of 1855 bp (SEQ ID NO:5). The nucleotide sequence includes an open reading frame (ORF) encoding a polypeptide of 405 amino acid residues (SEQ ID NO:6) with a predicted molecular weight of 44750 Daltons. The start codon is located at nucleotides 154-156 and the stop codon is located at nucleotides 1369-1371. The protein encoded by clone 11696905-0-47 is predicted by the PSORT program to localize extracellularly with a certainty of 0.7332. The program SignalP predicts that there is a signal peptide with the most probable cleavage site located between residues 25 and 26, in the sequence AQG-GP.

Real-time gene expression analysis was performed on SECP3 (clone 11696905-0-47). The results demonstrate that RNA sequences homologous to clone 11696905-0-47 are detected in various cell types. Cell types include adipose, adrenal gland, thyroid, brain, heart, skeletal

muscle, bone marrow, colon, bladder, liver, lung, mammary gland, placenta, and testis, and in neoplastic cells derived from renal carcinoma A498, lung carcinoma NCI-H460, and melanoma SK-MEL-28.

Accordingly, SECP3 nucleic acids according to the invention can be used to identify one or more of these cell types. The presence of RNA sequences homologous to a SECP3 nucleic in a sample indicates that the sample contains one or more of the above-cell types.

A search of the sequence databases using BLASTX reveals that clone 11696905-0-47 has 403 of 405 residues (99%) identical to, and 404 of 405 residues (99%) positive with, the 405 residue human angiopoietin-related protein (SPTREMBL-ACC:Q9Y5B3). Angiopoietin homologues are useful to stimulate cell growth and tissue development. The polypeptides of clone 11696905-0-47 tend to be found as multimeric proteins (see Example 7) and are believed to have angiogenic or hematopoietic activity. They can thus be used in assays for angiogenic activity, as well as used therapeutically to stimulate restoration of vascular structure in various tissues. Examples of such uses include, but are not limited to, treatment of full-thickness skin wounds, including venous stasis ulcers and other chronic, non-healing wounds, as well as fracture repair, skin grafting, reconstructive surgery, and establishment of vascular networks in transplanted cells and tissues.

The proteins of the invention encoded by clone 11696905-0-47 include the protein disclosed as being encoded by the ORF described herein, as well as any mature protein arising therefrom as a result of post-translational modifications. Thus, the proteins of the invention encompass both a precursor and any active forms of the clone 11696905-0-47 protein.

SECP4

A SECP4 nucleic acid and polypeptide according to the invention includes the nucleic acid sequence (SEQ ID NO:7) and encoded polypeptide sequence (SEQ ID NO:8) of 14578444.0.143. FIG. 4 illustrates the nucleic acid sequence and amino acid sequence, as well as the alignment between these two sequences.

Clone 14578444.0.143 was obtained from fetal brain. This clone includes a nucleotide sequence (SEQ ID NO:7) of 3026 bp. The nucleotide sequence includes an open reading frame (ORF) encoding a polypeptide of 776 amino acid residues (SEQ ID NO:8) with a predicted molecular weight of 86220.8 Daltons. The start codon is located at nucleotides 55-57 and the stop codon is located at nucleotides 2384-2386. The protein encoded by clone 14578444.0.143 is predicted by the PSORT program to localize in the endoplasmic reticulum (membrane) with a

certainty of 0.8200. The program SignalP predicts that there is a signal peptide with the most probable cleavage site located between residues 23 and 24 in the sequence AEA-RE.

A search of the sequence databases using BLASTX reveals that clone 14578444.0.143 has 655 of 757 residues (86%) identical to, and 702 of 757 residues (92%) positive with, the 956 residue murine matrilin-2 precursor protein (SWISSPROT-ACC:O08746), extending over residues 1-754 of the reference protein. Additional similarities are found with lower identities in residues 649-837 of the murine protein. Additionally, the search shows that there is a lower degree of similarity to murine matrilin-4 precursor. The protein of clone 14578444.0.143 also has 595 of 606 residues (98%) identical to, and 598 of 606 residues (98%) positive with, the 632 residue human matrilin-3 (PCT publication WO9904002-A1).

The matrilin proteins and polynucleotides can be used for treating a variety of developmental disorders (*e.g.*, renal tubular acidosis, anemia, Cushing's syndrome). The proteins can serve as targets for antagonists that should be of use in treating diseases related to abnormal vesicle trafficking. These may include, but are not limited to, diseases such as cystic fibrosis, glucose-galactose malabsorption syndrome, hypercholesterolaemia, diabetes mellitus, diabetes insipidus, hyper- and hypoglycemia, Graves disease, goiter, Cushing's disease, Addison's disease, gastrointestinal disorders including ulcerative colitis, gastric and duodenal ulcers, and other conditions associated with abnormal vesicle trafficking including AIDS, and allergies including hay fever, asthma, and urticaria (hives), autoimmune hemolytic anemia, proliferative glomerulonephritis, inflammatory bowel disease, multiple sclerosis, myasthenia gravis, rheumatoid and osteoarthritis, scleroderma, Chediak-Higashi and Sjogren's syndromes, systemic lupus erythematosus, toxic shock syndrome, traumatic tissue damage, and viral, bacterial, fungal, helminth, protozoal infections, a neoplastic disorder (*e.g.*, adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and cancers), or an immune disorder, (*e.g.*, AIDS, Addison's disease, adult respiratory distress syndrome, allergies, anemia, asthma, atherosclerosis, bronchitis, cholecystitis, Crohn's disease and ulcerative colitis).

The proteins of the invention encoded by clone 14578444.0.143 include the protein disclosed as being encoded by the ORF described herein, as well as any mature protein arising therefrom as a result of post-translational modifications. Thus, the proteins of the invention encompass both a precursor and any active forms of the proteins encoded by clone 14578444.0.143 (SECP4).

SECP5

A SECP5 nucleic acid and polypeptide according to the invention includes the nucleic acid sequence (SEQ ID NO:9) and encoded polypeptide sequence (SEQ ID NO:10) of clone 14578444.0.47. FIG. 5 illustrates the nucleic acid sequence and amino acid sequence, as well as the alignment between these two sequences.

Clone 14578444.0.47 was obtained from fetal brain. This clone includes a nucleotide sequence (SEQ ID NO:9) of 3447 bp. The nucleotide sequence includes an open reading frame (ORF) encoding a polypeptide of 959 amino acid residues (SEQ ID NO:10) with a predicted molecular weight of 107144 Daltons. The start codon is located at nucleotides 55-57 and the stop codon is located at nucleotides 2933-2935. The protein encoded by clone 14578444.0.47 is predicted by the PSORT program to localize to the endoplasmic reticulum (membrane) with a certainty of 0.8200. The program SignalP predicts that there is a signal peptide with the most probable cleavage site located between residues 23 and 24 in the sequence AEA-RE.

A search of the sequence databases using BLASTX reveals that clone 14578444.0.47 has 829 of 959 residues (86%) identical to, and 887 of 959 residues (92%) positive with, the 956 residue murine matrilin-2 precursor protein (ACC: SWISSPROT-ACC:O08746). The protein encoded by clone 14578444.0.47 also has 594 of 606 residues (98%) identical to, and 597 of 606 residues (98%) positive with, the 632 residue human matrilin-3 (PCT publication WO9904002). In addition, the protein encoded by clone 14578444.0.47 also has 616 of 678 residues (90%) identical to, and 632 of 678 residues (93%) positive with the 915 residue human protein PRO219 (PCT publication WO9914328-A2).

The proteins encoded by clones 14578444.0.143 (SECP4) and 14578444.0.47 (SECP5) are compared in an amino acid residue alignment shown in FIG. 11. It can be seen that the main portion of the two proteins starting with their amino-termini are virtually identical, and that short sequences in each corresponding to the carboxyl-terminal sequence of the shorter protein, clone 14578444.0.143, differ from one another. Furthermore, clone 14578444.0.47 has an extended carboxyl-terminal sequence that is missing in clone 14578444.0.143. Therefore, clones 14578444.0.143 (SECP4) and 14578444.0.47 (SECP5) are apparently related to one another as splice variants, with respect to their sequences at the carboxyl-terminal ends.

The matrilin proteins and polynucleotides can be used for treating a variety of developmental disorders (*e.g.*, renal tubular acidosis, anemia, Cushing's syndrome). The proteins can serve as targets for antagonists that should be of use in treating diseases related to abnormal

vesicle trafficking. These may include, but are not limited to, diseases such as cystic fibrosis, glucose-galactose malabsorption syndrome, hypercholesterolaemia, diabetes mellitus, diabetes insipidus, hyper- and hypoglycemia, Graves disease, goiter, Cushing's disease, Addison's disease, gastrointestinal disorders including ulcerative colitis, gastric and duodenal ulcers, and other conditions associated with abnormal vesicle trafficking including AIDS, and allergies including hay fever, asthma, and urticaria (hives), autoimmune hemolytic anemia, proliferative glomerulonephritis, inflammatory bowel disease, multiple sclerosis, myasthenia gravis, rheumatoid and osteoarthritis, scleroderma, Chediak-Higashi and Sjogren's syndromes, systemic lupus erythematosus, toxic shock syndrome, traumatic tissue damage, and viral, bacterial, fungal, helminth, protozoal infections, a neoplastic disorder (*e.g.,* adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and cancers), or an immune disorder, (*e.g.,* AIDS, Addison's disease, adult respiratory distress syndrome, allergies, anemia, asthma, atherosclerosis, bronchitis, cholecystitis, Crohn's disease and ulcerative colitis).

The proteins of the invention encoded by clone 14578444.0.47 include the protein disclosed as being encoded by the ORF described herein, as well as any mature protein arising therefrom as a result of post-translational modifications. Thus, the proteins of the invention encompass both a precursor and any active forms of the proteins encoded by clone 14578444.0.47 (SECP5).

SECP6

A SECP6 nucleic acid and polypeptide according to the invention includes the nucleic acid sequence (SEQ ID NO:11) and encoded polypeptide sequence (SEQ ID NO:12) of clone 14998905.0.65. FIG. 6 illustrates the nucleic acid sequence and amino acid sequence, as well as the alignment between these two sequences.

Clone 14998905.0.65 was obtained from lymphoid tissue, in particular, from the lymph node. This clone includes a nucleotide sequence (SEQ ID NO:11) of 967 bp. The nucleotide sequence includes an open reading frame (ORF) encoding a polypeptide of 245 amino acid residues (SEQ ID NO:12) with a predicted molecular weight of 27327.2 Daltons. The start codon is located at nucleotides 166-168 and the stop codon is located at nucleotides 902-904. The protein encoded by clone 14998905.0.65 is predicted by the PSORT program to localize in the microbody (peroxisome) with a certainty of 0.7480. PSORT predicts that there is no amino-terminal signal sequence. Conversely, the program SignalP predicts that there is a signal peptide

with the most probable cleavage site located between residues 20 and 21, in the sequence GIG-AE.

A search of the sequence databases using BLASTX reveals that clone 14998905.0.65 has 204 of 226 residues (90%) identical to, and 214 of 226 residues (94%) positive with, the 834 residue murine semaphorin 4C precursor protein (SWISSPROT-ACC:Q64151). Semaphorin 4C is indicated as being a Type I membrane protein widely expressed in the nervous system during development. In addition, it contains one immunoglobulin-like C2-type domain. The protein encoded by clone 14998905.0.65 also has similarities to mouse CD100 antigen (PCT publication WO9717368-A1) and to human semaphorin (JP10155490-A).

The proteins of the invention encoded by clone 14998905.0.65 include the protein disclosed as being encoded by the ORF described herein, as well as any mature protein arising therefrom as a result of post-translational modifications. Thus, the proteins of the invention encompass both a precursor and any active forms of the clone 14998905.0.65 protein.

SECP7

A SECP7 nucleic acid and polypeptide according to the invention includes the nucleic acid sequence (SEQ ID NO:13) and encoded polypeptide sequence (SEQ ID NO:14) of clone 16406477.0.206. FIG. 7 illustrates the nucleic acid sequence and amino acid sequence, as well as the alignment between these two sequences.

Clone 16406477.0.206 was obtained from testis. In addition, sequences of clone 16406477.0.206 were also found in an RNA pool derived from adrenal gland, mammary gland, prostate gland, testis, uterus, bone marrow, melanoma, pituitary gland, thyroid gland and spleen. This clone includes a nucleotide sequence (SEQ ID NO:13) comprising of 1359 bp with an open reading frame (ORF) encoding a polypeptide of 385 amino acid residues (SEQ ID NO:14) with a predicted molecular weight of 43087.3 Daltons. The start codon is located at nucleotides 45-47 and the stop codon is located at nucleotides 1201-1203. The protein encoded by clone 16406477.0.206 is predicted by the PSORT program to localize extracellularly with a certainty of 0.5804 and to have a cleavable amino-terminal signal sequence. The program SignalP predicts that there is a signal peptide with the most probable cleavage site located between residues 39 and 40, in the sequence CWG-AG.

Real-time expression analysis was performed on SECP7 (clone 16406477.0.206). The results demonstrate that RNA homologous to this clone is found in multiple cell and tissue types. These cells and tissues include brain, mammary gland, and testis, and in neoplastic cells derived

from ovarian carcinoma OVCAR-3, ovarian carcinoma OVCAR-5, ovarian carcinoma OVCAR-8, ovarian carcinoma IGROV-1, breast carcinoma (pleural effusion) T47D, breast carcinoma BT-549, melanoma M14. Real-time gene expression analysis was performed on SECP3 (clone 11696905-0-47). The results demonstrate that RNA sequences homologous to clone 11696905-0-47 are detected in various cell types. Cell types include adipose, adrenal gland, thyroid, brain, heart, skeletal muscle, bone marrow, colon, bladder, liver, lung, mammary gland, placenta, and testis, and in neoplastic cells derived from renal carcinoma A498, lung carcinoma NCI-H460, and melanoma SK-MEL-28.

Accordingly, SECP7 nucleic acids according to the invention can be used to identify one or more of these cell types. The presence of RNA sequences homologous to a SECP7 nucleic in a sample indicates that the sample contains one or more of the above-cell types.

A search of the sequence databases using BLASTX reveals that clone 16406477.0.206 is 100% identical to a human testis-specific protein TSP50 (SPTREMBL-ACC:Q9UI38) with a trypsin/chymotrypsin-like domain. In addition, the protein encoded by clone 16406477.0.206 has low similarity to the 343 residue human prostasin precursor (EC 3.4.21.-) (SWISSPROT ACC:Q16651).

The proteins of the invention encoded by clone 16406477.0.206 include the protein disclosed as being encoded by the ORF described herein, as well as any mature protein arising therefrom as a result of post-translational modifications. Thus, the proteins of the invention encompass both a precursor and any active forms of the clone 16406477.0.206 protein.

SECP8

A SECP8 nucleic acid and polypeptide according to the invention includes the nucleic acid sequence (SEQ ID NO:15) and encoded polypeptide sequence (SEQ ID NO:16) of clone 11618130.0.184. FIG. 8 illustrates the nucleic acid sequence and amino acid sequence, as well as the alignment between these two sequences.

Clone 11618130.0.184 includes a nucleotide sequence (SEQ ID NO:15) of 1445 bp. The nucleotide sequence includes an open reading frame (ORF) encoding a polypeptide of 198 amino acid residues (SEQ ID NO:16) with a predicted molecular weight of 20659 Daltons. The start codon is located at nucleotides 732-734 and the stop codon is located at nucleotides 1326-1328. The protein encoded by clone 11618130.0.184 is predicted by the PSORT program to localize in the cytoplasm. The program SignalP predicts that there is no signal peptide.

Clones 11618130.0.184 (SECP8) and 11618130.0.27 (SECP2) resemble each other in that they are identical over most of their common sequences, and differ only at the carboxyl-terminal end. In addition, clone 11618130.0.27 extends further at the carboxyl-terminal end than does clone 11618130.0.184. An alignment of clones 11618130.0.27 and 11618130.0.184 is shown in FIG. 10.

The proteins of the invention encoded by clone 11618130.0.184 include the protein disclosed as being encoded by the ORF described herein, as well as any mature protein arising therefrom as a result of post-translational modifications. Thus, the proteins of the invention encompass both a precursor and any active forms of the 11618130.0.184 protein.

SECP9

A SECP9 nucleic acid and polypeptide according to the invention includes the nucleic acid sequence (SEQ ID NO:17) and encoded polypeptide sequence (SEQ ID NO:18) of clone 21637262.0.64. FIG. 9 illustrates the nucleic acid sequence and amino acid sequence, as well as the alignment between these two sequences.

Clone 21637262.0.64 was obtained from salivary gland. This clone includes a nucleotide sequence (SEQ ID NO:17) of 1600 bp. The nucleotide sequence includes an open reading frame (ORF) encoding a polypeptide of435 amino acid residues (SEQ ID NO:18) with a predicted molecular weight of 47162.5 Daltons. The start codon is located at nucleotides 51-53 and the stop codon is located at nucleotides 1356-1358. The protein encoded by clone 21637262.0.64 is predicted by the PSORT program to localize in the cytoplasm with a certainty of 0.4500. The program PSORT and program SignalP predict that the protein appears to have no amino-terminal signal sequence.

Real-time expression analysis was performed on SECP9 (clone 21637262.0.64). The results demonstrate that RNA homologous to this clone is present in multiple tissue and cell types. The relative amounts of RNA in various cell types are shown in FIG. 14 (see also the Examples, below). The cells include myometrium, placenta, uterus, prostate, and testis, and neoplastic cells derived from breast carcinoma (pleural effusion) T47D, breast carcinoma (pleural effusion) MDA-MB-231, breast carcinoma BT-549, ovarian carcinoma OVCAR-3, ovarian carcinoma OVCAR-5, prostate carcinoma (bone metastases) PC-3, melanoma M14, and melanoma LOX IMVI.

Accordingly, SECP9 nucleic acids according to the invention can be used to identify one or more of these cell types. The presence of RNA sequences homologous to a SECP9 nucleic in a sample indicates that the sample contains one or more of the above-cell types.

A search of the sequence databases using BLASTX reveals that clone 21637262.0.64 has 23 of 420 residues (29%) identical to, and 201 of 420 residues (47%) positive with, the 1130 residue murine protein repetin (SWISSPROT-ACC:P97347). Repetin is a member of the "fused gene" subgroup within the S100 gene family that is an epidermal differentiation protein.

The proteins of the invention encoded by clone 21637262.0.64 include the protein disclosed as being encoded by the ORF described herein, as well as any mature protein arising therefrom as a result of post-translational modifications. Thus, the proteins of the invention encompass both a precursor and any active forms of the clone 21637262.0.64 protein.

SECP10

A SECP10 nucleic acid and polypeptide according to the invention includes the nucleic acid sequence (SEQ ID NO:40 and encoded polypeptide sequence (SEQ ID NO:41) of clone CG106318. FIG. 15 illustrates the nucleic acid sequence and amino acid sequences. This clone includes a nucleotide sequence (SEQ ID NO:40) of 4810 bp. The nucleotide sequence includes an open reading frame (ORF) encoding a polypeptide of 1588 amino acid residues (SEQ ID NO:41). The start codon is located at nucleotides 18-21 and the stop codon is located at nucleotides 4782-4785. The protein encoded by clone CG106318-01 is predicted by the PSORT program to localize in the nucleus with a certainty of 0.3500. The program PSORT and program SignalP predict that the protein appears to have no amino-terminal signal sequence.

Real-time expression analysis was performed on SECP10 (clone CG106318). The results demonstrate that RNA homologous to this clone is present in multiple tissue and cell types.

Accordingly, SECP10 nucleic acids according to the invention can be used to identify one or more of these tissue types. The presence of RNA sequences homologous to a SECP10 nucleic acid in a sample indicates that the sample contains one or more of the above-tissue types.

A search of the sequence databases using BLASTX reveals that clone CG106318 has 1587 out of 1588 (99.9%) of its residues identical to a human protein utilized in the treatment of central nervous system disorders (AAM39295 to HYSEQ INC.).

The proteins of the invention encoded by clone CG106318-01 include the protein disclosed as being encoded by the ORF described herein, as well as any mature protein arising

therefrom as a result of post-translational modifications. Thus, the proteins of the invention encompass both a precursor and any active forms of the clone CG106318-01 protein.

PSORT --- Prediction of Protein Translocation Sites version 5.8

Results Summary:

```
                plasma membrane --- Certainty=0.7000(Affirmative) < succ>
                        nucleus --- Certainty=0.3500(Affirmative) < succ>
          microbody (peroxisome) --- Certainty=0.3000(Affirmative) < succ>
endoplasmic reticulum (membrane) --- Certainty=0.2000(Affirmative) < succ>
```

PFAM Domain Analysis
Query: 106318-01

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
tsp_1	Thrombospondin type 1 domain	169.5	5.4e-47	11
toxin	Snake toxin	-16.1	1.3	1
DUF18	Domain of unknown function DUF18	-55.9	7.8	1
Keratin_B2	Keratin, high sulfur B2 protein	-81.1	6.6	1

Sequences producing High-scoring Segment Pairs:	Score	P(N)	N
gb:GENBANK-ID:AX079870|acc:AX079870.1 Sequence 1 from Pat.....	24050	0.0	1
gb:GENBANK-ID:AB023177|acc:AB023177.1 Homo sapiens mRNA f....	19495	0.0	1
gb:GENBANK-ID:AB051466|acc:AB051466.1 Homo sapiens mRNA f... ...	3611	5.3e-269	6
gb:GENBANK-ID:AB006087|acc:AB006087.1 Danio rerio mRNA fo.....	272	0.16	1
gb:GENBANK-ID:AF111298|acc:AF111298.1 HIV-1 isolate eur-0.....	185	0.998	1

BLASTP: (1588 letters)

Database: Non-Redundant Composite Protein
704,847 sequences; 219,724,008 total letters.
Searching....10....20....30....40....50....60....70....80....90....100% done

Sequences producing High-scoring Segment Pairs:	Smallest Sum High Probability Score	P(N)	N
ptnr:REMTREMBL-ACC:CAC32422 Sequence 1 from Patent WO0105...	8965	0.0	1
ptnr:SPTREMBL-ACC:Q9UPZ6 KIAA0960 PROTEIN - Homo sapiens ...	7298	0.0	1
ptnr:SPTREMBL-ACC:Q9C0I4 KIAA1679 PROTEIN - Homo sapiens ...	3983	0.0	1
ptnr:SPTREMBL-ACC:O60407 HYPOTHETICAL PROTEIN - Homo sapi...	3026	3.1e-315	1

TABLE 2. BLASTN VERSUS GENBANK COMPOSITE

Sequences producing High-scoring Segment Pairs:	Score	P(N)	N
gb:GENBANK-ID:AX079870|acc:AX079870.1 Sequence 1 from Pat.....	24050	0.0	1
gb:GENBANK-ID:AB023177|acc:AB023177.1 Homo sapiens mRNA f....	19495	0.0	1
gb:GENBANK-ID:AB051466|acc:AB051466.1 Homo sapiens mRNA f.....	3611	5.3e-269	6
gb:GENBANK-ID:AB006087|acc:AB006087.1 Danio rerio mRNA fo.....	272	0.16	1
gb:GENBANK-ID:AF111298|acc:AF111298.1 HIV-1 isolate eur-0.....	185	0.998	1

```
>gb:GENBANK-ID:AX079870|acc:AX079870.1 Sequence 1 from Patent WO0105971 - Homo
          sapiens, 6373 bp.(SEQ ID NO:58)
          Length = 6373

 Plus Strand HSPs:

Score = 24050 (3608.5 bits), Expect = 0.0, P = 0.0
Identities = 4810/4810 (100%), Positives = 4810/4810 (100%), Strand = Plus / Plus
```

```
Query:    1 GTCCATGGGGCCGATGTATGGGAGATGAATGTGGTCCCGGAGGCATCCAAACGAGGGCTG 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  218 GTCCATGGGGCCGATGTATGGGAGATGAATGTGGTCCCGGAGGCATCCAAACGAGGGCTG 277

Query:   61 TGTGGTGTGCTCATGTGGAGGGATGGACTACACTGCATACTAACTGTAAGCAGGCCGAGA 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  278 TGTGGTGTGCTCATGTGGAGGGATGGACTACACTGCATACTAACTGTAAGCAGGCCGAGA 337

Query:  121 GACCCAATAACCAGCAGAATTGTTTCAAAGTTTGCGATTGGCACAAAGAGTTGTACGACT 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  338 GACCCAATAACCAGCAGAATTGTTTCAAAGTTTGCGATTGGCACAAAGAGTTGTACGACT 397

Query:  181 GGAGACTGGGACCTTGGAATCAGTGTCAGCCCGTGATTTCAAAAAGCCTAGAGAAACCTC 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  398 GGAGACTGGGACCTTGGAATCAGTGTCAGCCCGTGATTTCAAAAAGCCTAGAGAAACCTC 457

Query:  241 TTGAGTGCATTAAGGGGGAAGAAGGTATTCAGGTGAGGGAGATAGCGTGCATCCAGAAAG 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  458 TTGAGTGCATTAAGGGGGAAGAAGGTATTCAGGTGAGGGAGATAGCGTGCATCCAGAAAG 517

Query:  301 ACAAAGACATTCCTGCGGAGGATATCATCTGTGAGTACTTTGAGCCCAAGCCTCTCCTGG 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  518 ACAAAGACATTCCTGCGGAGGATATCATCTGTGAGTACTTTGAGCCCAAGCCTCTCCTGG 577

Query:  361 AGCAGGCTTGCCTCATTCCTTGCCAGCAAGATTGCATCGTGTCTGAATTTTCTGCCTGGT 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  578 AGCAGGCTTGCCTCATTCCTTGCCAGCAAGATTGCATCGTGTCTGAATTTTCTGCCTGGT 637

Query:  421 CCGAATGCTCCAAGACCTGCGGCAGCGGGCTCCAGCACCGGACGCGTCATGTGGTGGCGC 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  638 CCGAATGCTCCAAGACCTGCGGCAGCGGGCTCCAGCACCGGACGCGTCATGTGGTGGCGC 697

Query:  481 CCCCGCAGTTCGGAGGCTCTGGCTGTCCAAACCTGACGGAGTTCCAGGTGTGCCAATCCA 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  698 CCCCGCAGTTCGGAGGCTCTGGCTGTCCAAACCTGACGGAGTTCCAGGTGTGCCAATCCA 757

Query:  541 GTCCATGCGAGGCCGAGGAGCTCAGGTACAGCCTGCATGTGGGGCCCTGGAGCACCTGCT 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  758 GTCCATGCGAGGCCGAGGAGCTCAGGTACAGCCTGCATGTGGGGCCCTGGAGCACCTGCT 817

Query:  601 CAATGCCCCACTCCCGACAAGTAAGACAAGCAAGGAGACGCGGGAAGAATAAAGAACGGG 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  818 CAATGCCCCACTCCCGACAAGTAAGACAAGCAAGGAGACGCGGGAAGAATAAAGAACGGG 877

Query:  661 AAAAGGACCGCAGCAAAGGAGTAAAGGATCCAGAAGCCCGCGAGCTTATTAAGAAAAAGA 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  878 AAAAGGACCGCAGCAAAGGAGTAAAGGATCCAGAAGCCCGCGAGCTTATTAAGAAAAAGA 937

Query:  721 GAAACAGAAACAGGCAGAACAGACAAGAGAACAAATATTGGGACATCCAGATTGGATATC 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  938 GAAACAGAAACAGGCAGAACAGACAAGAGAACAAATATTGGGACATCCAGATTGGATATC 997

Query:  781 AGACCAGAGAGGTTATGTGCATTAACAAGACGGGGAAAGCTGCTGATTTAAGCTTTTGCC 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  998 AGACCAGAGAGGTTATGTGCATTAACAAGACGGGGAAAGCTGCTGATTTAAGCTTTTGCC 1057

Query:  841 AGCAAGAGAAGCTTCCAATGACCTTCCAGTCCTGTGTGATCACCAAAGAGTGCCAGGTTT 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1058 AGCAAGAGAAGCTTCCAATGACCTTCCAGTCCTGTGTGATCACCAAAGAGTGCCAGGTTT 1117

Query:  901 CCGAGTGGTCAGAGTGGAGCCCCTGCTCAAAAACATGCCATGACATGGTGTCCCCTGCAG 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1118 CCGAGTGGTCAGAGTGGAGCCCCTGCTCAAAAACATGCCATGACATGGTGTCCCCTGCAG 1177

Query:  961 GCACTCGTGTAAGGACACGAACCATCAGGCAGTTTCCCATTGGCAGTGAAAAGGAGTGTC 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1178 GCACTCGTGTAAGGACACGAACCATCAGGCAGTTTCCCATTGGCAGTGAAAAGGAGTGTC 1237

Query: 1021 CAGAATTTGAAGAAAAAGAACCCTGTTTGTCTCAAGGAGATGGAGTTGTCCCCTGTGCCA 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1238 CAGAATTTGAAGAAAAAGAACCCTGTTTGTCTCAAGGAGATGGAGTTGTCCCCTGTGCCA 1297

Query: 1081 CGTATGGCTGGAGAACTACAGAGTGGACTGAGTGCCGTGTGGACCCTTTGCTCAGTCAGC 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1298 CGTATGGCTGGAGAACTACAGAGTGGACTGAGTGCCGTGTGGACCCTTTGCTCAGTCAGC 1357
```

```
Query:  1141 AGGACAAGAGGCGCGGCAACCAGACGGCCCTCTGTGGAGGGGGCATCCAGACCCGAGAGG 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1358 AGGACAAGAGGCGCGGCAACCAGACGGCCCTCTGTGGAGGGGGCATCCAGACCCGAGAGG 1417

Query:  1201 TGTACTGCGTGCAGGCCAACGAAAACCTCCTCTCACAATTAAGTACCCACAAGAACAAAG 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1418 TGTACTGCGTGCAGGCCAACGAAAACCTCCTCTCACAATTAAGTACCCACAAGAACAAAG 1477

Query:  1261 AAGCCTCAAAGCCAATGGACTTAAAATTATGCACTGGACCTATCCCTAATACTACACAGC 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1478 AAGCCTCAAAGCCAATGGACTTAAAATTATGCACTGGACCTATCCCTAATACTACACAGC 1537

Query:  1321 TGTGCCACATTCCTTGTCCAACTGAATGTGAAGTTTCACCTTGGTCAGCTTGGGGACCTT 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1538 TGTGCCACATTCCTTGTCCAACTGAATGTGAAGTTTCACCTTGGTCAGCTTGGGGACCTT 1597

Query:  1381 GTACTTATGAAAACTGTAATGATCAGCAAGGGAAAAAAGGCTTCAAACTGAGGAAGCGGC 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1598 GTACTTATGAAAACTGTAATGATCAGCAAGGGAAAAAAGGCTTCAAACTGAGGAAGCGGC 1657

Query:  1441 GCATTACCAATGAGCCCACTGGAGGCTCTGGGGTAACCGGAAACTGCCCTCACTTACTGG 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1658 GCATTACCAATGAGCCCACTGGAGGCTCTGGGGTAACCGGAAACTGCCCTCACTTACTGG 1717

Query:  1501 AAGCCATTCCCTGTGAAGAGCCTGCCTGTTATGACTGGAAAGCGGTGAGACTGGGAGACT 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1718 AAGCCATTCCCTGTGAAGAGCCTGCCTGTTATGACTGGAAAGCGGTGAGACTGGGAGACT 1777

Query:  1561 GCGAGCCAGATAACGGAAAGGAGTGTGGTCCAGGCACGCAAGTTCAAGAGGTTGTGTGCA 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1778 GCGAGCCAGATAACGGAAAGGAGTGTGGTCCAGGCACGCAAGTTCAAGAGGTTGTGTGCA 1837

Query:  1621 TCAACAGTGATGGAGAAGAAGTTGACAGACAGCTGTGCAGAGATGCCATCTTCCCCATCC 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1838 TCAACAGTGATGGAGAAGAAGTTGACAGACAGCTGTGCAGAGATGCCATCTTCCCCATCC 1897

Query:  1681 CTGTGGCCTGTGATGCCCCATGCCCGAAAGACTGTGTGCTCAGCACATGGTCTACGTGGT 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1898 CTGTGGCCTGTGATGCCCCATGCCCGAAAGACTGTGTGCTCAGCACATGGTCTACGTGGT 1957

Query:  1741 CCTCCTGCTCACACACCTGCTCAGGGAAAACGACAGAAGGGAAACAGATACGAGCACGAT 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1958 CCTCCTGCTCACACACCTGCTCAGGGAAAACGACAGAAGGGAAACAGATACGAGCACGAT 2017

Query:  1801 CCATTCTGGCCTATGCGGGTGAAGAAGGTGGAATTCGCTGTCCAAATAGCAGTGCTTTGC 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  2018 CCATTCTGGCCTATGCGGGTGAAGAAGGTGGAATTCGCTGTCCAAATAGCAGTGCTTTGC 2077

Query:  1861 AAGAAGTACGAAGCTGTAATGAGCATCCTTGCACAGTGTACCACTGGCAAACTGGTCCCT 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  2078 AAGAAGTACGAAGCTGTAATGAGCATCCTTGCACAGTGTACCACTGGCAAACTGGTCCCT 2137

Query:  1921 GGGGCCAGTGCATTGAGGACACCTCAGTATCGTCCTTCAACACAACTACGACTTGGAATG 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  2138 GGGGCCAGTGCATTGAGGACACCTCAGTATCGTCCTTCAACACAACTACGACTTGGAATG 2197

Query:  1981 GGGAGGCCTCCTGCTCTGTCGGCATGCAGACAAGAAAAGTCATCTGTGTGCGAGTCAATG 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  2198 GGGAGGCCTCCTGCTCTGTCGGCATGCAGACAAGAAAAGTCATCTGTGTGCGAGTCAATG 2257

Query:  2041 TGGGCCAAGTGGGACCCAAAAAATGTCCTGAAAGCCTTCGACCTGAAACTGTAAGGCCTT 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  2258 TGGGCCAAGTGGGACCCAAAAAATGTCCTGAAAGCCTTCGACCTGAAACTGTAAGGCCTT 2317

Query:  2101 GTCTGCTTCCTTGTAAGAAGGACTGTATTGTGACCCCATATAGTGACTGGACATCATGCC 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  2318 GTCTGCTTCCTTGTAAGAAGGACTGTATTGTGACCCCATATAGTGACTGGACATCATGCC 2377

Query:  2161 CCTCTTCGTGTAAAGAAGGGGACTCCAGTATCAGGAAGCAGTCTAGGCATCGGGTCATCA 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  2378 CCTCTTCGTGTAAAGAAGGGGACTCCAGTATCAGGAAGCAGTCTAGGCATCGGGTCATCA 2437

Query:  2221 TTCAGCTGCCAGCCAACGGGGGCCGAGACTGCACAGATCCCCTCTATGAAGAGAAGGCCT 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  2438 TTCAGCTGCCAGCCAACGGGGGCCGAGACTGCACAGATCCCCTCTATGAAGAGAAGGCCT 2497
```

```
Query: 2281 GTGAGGCACCTCAAGCGTGCCAAAGCTACAGGTGGAAGACTCACAAATGGCGCAGATGCC 2340
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2498 GTGAGGCACCTCAAGCGTGCCAAAGCTACAGGTGGAAGACTCACAAATGGCGCAGATGCC 2557

Query: 2341 AATTAGTCCCTTGGAGCGTGCAACAAGACAGCCCTGGAGCACAGGAAGGCTGTGGGCCTG 2400
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2558 AATTAGTCCCTTGGAGCGTGCAACAAGACAGCCCTGGAGCACAGGAAGGCTGTGGGCCTG 2617

Query: 2401 GGCGACAGGCAAGAGCCATTACTTGTCGCAAGCAAGATGGAGGACAGGCTGGAATCCATG 2460
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2618 GGCGACAGGCAAGAGCCATTACTTGTCGCAAGCAAGATGGAGGACAGGCTGGAATCCATG 2677

Query: 2461 AGTGCCTACAGTATGCAGGCCCTGTGCCAGCCCTTACCCAGGCCTGCCAGATCCCCTGCC 2520
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2678 AGTGCCTACAGTATGCAGGCCCTGTGCCAGCCCTTACCCAGGCCTGCCAGATCCCCTGCC 2737

Query: 2521 AGGATGACTGTCAATTGACCAGCTGGTCCAAGTTTTCTTCATGCAATGGAGACTGTGGTG 2580
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2738 AGGATGACTGTCAATTGACCAGCTGGTCCAAGTTTTCTTCATGCAATGGAGACTGTGGTG 2797

Query: 2581 CAGTTAGGACCAGAAAGCGCACTCTTGTTGGAAAAAGTAAAAAGAAGGAAAAATGTAAAA 2640
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2798 CAGTTAGGACCAGAAAGCGCACTCTTGTTGGAAAAAGTAAAAAGAAGGAAAAATGTAAAA 2857

Query: 2641 ATTCCCATTTGTATCCCCTGATTGAGACTCAGTATTGTCCTTGTGACAAATATAATGCAC 2700
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2858 ATTCCCATTTGTATCCCCTGATTGAGACTCAGTATTGTCCTTGTGACAAATATAATGCAC 2917

Query: 2701 AACCTGTGGGGAACTGGTCAGACTGTATTTTACCAGAGGGAAAAGTGGAAGTGTTGCTGG 2760
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2918 AACCTGTGGGGAACTGGTCAGACTGTATTTTACCAGAGGGAAAAGTGGAAGTGTTGCTGG 2977

Query: 2761 GAATGAAAGTACAAGGAGACATCAAGGAATGCGGACAAGGATATCGTTACCAAGCAATGG 2820
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2978 GAATGAAAGTACAAGGAGACATCAAGGAATGCGGACAAGGATATCGTTACCAAGCAATGG 3037

Query: 2821 CATGCTACGATCAAAATGGCAGGCTTGTGGAAACATCTAGATGTAACAGCCATGGTTACA 2880
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3038 CATGCTACGATCAAAATGGCAGGCTTGTGGAAACATCTAGATGTAACAGCCATGGTTACA 3097

Query: 2881 TTGAGGAGGCCTGCATCATCCCCTGCCCCTCAGACTGCAAGCTCAGTGAGTGGTCCAACT 2940
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3098 TTGAGGAGGCCTGCATCATCCCCTGCCCCTCAGACTGCAAGCTCAGTGAGTGGTCCAACT 3157

Query: 2941 GGTCGCGCTGCAGCAAGTCCTGTGGGAGTGGTGTGAAGGTTCGTTCTAAATGGCTGCGTG 3000
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3158 GGTCGCGCTGCAGCAAGTCCTGTGGGAGTGGTGTGAAGGTTCGTTCTAAATGGCTGCGTG 3217

Query: 3001 AAAAACCATATAATGGAGGAAGGCCTTGCCCCAAACTGGACCATGTCAACCAGGCACAGG 3060
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3218 AAAAACCATATAATGGAGGAAGGCCTTGCCCCAAACTGGACCATGTCAACCAGGCACAGG 3277

Query: 3061 TGTATGAGGTTGTCCCATGCCACAGTGACTGCAACCAGTACCTATGGGTCACAGAGCCCT 3120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3278 TGTATGAGGTTGTCCCATGCCACAGTGACTGCAACCAGTACCTATGGGTCACAGAGCCCT 3337

Query: 3121 GGAGCATCTGCAAGGTGACCTTTGTGAATATGCGGGAGAACTGTGGAGAGGGCGTGCAAA 3180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3338 GGAGCATCTGCAAGGTGACCTTTGTGAATATGCGGGAGAACTGTGGAGAGGGCGTGCAAA 3397

Query: 3181 CCCGAAAAGTGAGATGCATGCAGAATACAGCAGATGGCCCTTCTGAACATGTAGAGGATT 3240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3398 CCCGAAAAGTGAGATGCATGCAGAATACAGCAGATGGCCCTTCTGAACATGTAGAGGATT 3457

Query: 3241 ACCTCTGTGACCCAGAAGAGATGCCCCTGGGCTCTAGAGTGTGCAAATTACCATGCCCTG 3300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3458 ACCTCTGTGACCCAGAAGAGATGCCCCTGGGCTCTAGAGTGTGCAAATTACCATGCCCTG 3517

Query: 3301 AGGACTGTGTGATATCTGAATGGGGTCCATGGACCCAATGTGTTTTGCCTTGCAATCAAA 3360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3518 AGGACTGTGTGATATCTGAATGGGGTCCATGGACCCAATGTGTTTTGCCTTGCAATCAAA 3577

Query: 3361 GCAGTTTCCGGCAAAGGTCAGCTGATCCCATCAGACAACCAGCTGATGAAGGAAGATCTT 3420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3578 GCAGTTTCCGGCAAAGGTCAGCTGATCCCATCAGACAACCAGCTGATGAAGGAAGATCTT 3637
```

```
Query:  3421 GCCCTAATGCTGTTGAGAAAGAACCCTGTAACCTGAACAAAAACTGCTACCACTATGATT 3480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  3638 GCCCTAATGCTGTTGAGAAAGAACCCTGTAACCTGAACAAAAACTGCTACCACTATGATT 3697

Query:  3481 ATAATGTAACAGACTGGAGTACATGTCAGCTGAGTGAGAAGGCAGTTTGTGGAAATGGAA 3540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  3698 ATAATGTAACAGACTGGAGTACATGTCAGCTGAGTGAGAAGGCAGTTTGTGGAAATGGAA 3757

Query:  3541 TAAAAACAAGGATGTTGGATTGTGTTCGAAGTGATGGCAAGTCAGTTGACCTGAAATATT 3600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  3758 TAAAAACAAGGATGTTGGATTGTGTTCGAAGTGATGGCAAGTCAGTTGACCTGAAATATT 3817

Query:  3601 GTGAAGCGCTTGGCTTGGAGAAGAACTGGCAGATGAACACGTCCTGCATGGTGGAATGCC 3660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  3818 GTGAAGCGCTTGGCTTGGAGAAGAACTGGCAGATGAACACGTCCTGCATGGTGGAATGCC 3877

Query:  3661 CTGTGAACTGTCAGCTTTCTGATTGGTCTCCTTGGTCAGAATGTTCTCAAACATGTGGCC 3720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  3878 CTGTGAACTGTCAGCTTTCTGATTGGTCTCCTTGGTCAGAATGTTCTCAAACATGTGGCC 3937

Query:  3721 TCACAGGAAAAATGATCCGAAGACGAACAGTGACCCAGCCCTTTCAAGGTGATGGAAGAC 3780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  3938 TCACAGGAAAAATGATCCGAAGACGAACAGTGACCCAGCCCTTTCAAGGTGATGGAAGAC 3997

Query:  3781 CATGCCCTTCCCTGATGGACCAGTCCAAACCCTGCCCAGTGAAGCCTTGTTATCGGTGGC 3840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  3998 CATGCCCTTCCCTGATGGACCAGTCCAAACCCTGCCCAGTGAAGCCTTGTTATCGGTGGC 4057

Query:  3841 AATATGGCCAGTGGTCTCCATGCCAAGTGCAGGAGGCCCAGTGTGGAGAAGGGACCAGAA 3900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  4058 AATATGGCCAGTGGTCTCCATGCCAAGTGCAGGAGGCCCAGTGTGGAGAAGGGACCAGAA 4117

Query:  3901 CAAGGAACATTTCTTGTGTAGTAAGTGATGGGTCAGCTGATGATTTCAGCAAAGTGGTGG 3960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  4118 CAAGGAACATTTCTTGTGTAGTAAGTGATGGGTCAGCTGATGATTTCAGCAAAGTGGTGG 4177

Query:  3961 ATGAGGAATTCTGTGCTGACATTGAACTCATTATAGATGGTAATAAAAATATGGTTCTGG 4020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  4178 ATGAGGAATTCTGTGCTGACATTGAACTCATTATAGATGGTAATAAAAATATGGTTCTGG 4237

Query:  4021 AGGAATCCTGCAGCCAGCCTTGCCCAGGTGACTGTTATTTGAAGGACTGGTCTTCCTGGA 4080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  4238 AGGAATCCTGCAGCCAGCCTTGCCCAGGTGACTGTTATTTGAAGGACTGGTCTTCCTGGA 4297

Query:  4081 GCCTGTGTCAGCTGACCTGTGTGAATGGTGAGGATCTAGGCTTTGGTGGAATACAGGTCA 4140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  4298 GCCTGTGTCAGCTGACCTGTGTGAATGGTGAGGATCTAGGCTTTGGTGGAATACAGGTCA 4357

Query:  4141 GATCCAGACCGGTGATTATACAAGAACTAGAGAATCAGCATCTGTGCCCAGAGCAGATGT 4200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  4358 GATCCAGACCGGTGATTATACAAGAACTAGAGAATCAGCATCTGTGCCCAGAGCAGATGT 4417

Query:  4201 TAGAAACAAAATCATGTTATGATGGACAGTGCTATGAATATAAATGGATGGCCAGTGCTT 4260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  4418 TAGAAACAAAATCATGTTATGATGGACAGTGCTATGAATATAAATGGATGGCCAGTGCTT 4477

Query:  4261 GGAAGGGCTCTTCCCGAACAGTGTGGTGTCAAAGGTCAGATGGTATAAATGTAACAGGGG 4320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  4478 GGAAGGGCTCTTCCCGAACAGTGTGGTGTCAAAGGTCAGATGGTATAAATGTAACAGGGG 4537

Query:  4321 GCTGCTTGGTGATGAGCCAGCCTGATGCCGACAGGTCTTGTAACCCACCGTGTAGTCAAC 4380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  4538 GCTGCTTGGTGATGAGCCAGCCTGATGCCGACAGGTCTTGTAACCCACCGTGTAGTCAAC 4597

Query:  4381 CCCACTCGTACTGTAGCGAGACAAAAACATGCCATTGTGAAGAAGGGTACACTGAAGTCA 4440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  4598 CCCACTCGTACTGTAGCGAGACAAAAACATGCCATTGTGAAGAAGGGTACACTGAAGTCA 4657

Query:  4441 TGTCTTCTAACAGCACCCTTGAGCAATGCACACTTATCCCCGTGGTGGTATTACCCACCA 4500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  4658 TGTCTTCTAACAGCACCCTTGAGCAATGCACACTTATCCCCGTGGTGGTATTACCCACCA 4717

Query:  4501 TGGAGGACAAAAGAGGAGATGTGAAAACCAGTCGGGCTGTACATCCAACCCAACCCTCCA 4560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  4718 TGGAGGACAAAAGAGGAGATGTGAAAACCAGTCGGGCTGTACATCCAACCCAACCCTCCA 4777
```

```
Query:  4561 GTAACCCAGCAGGACGGGGAAGGACCTGGTTTCTACAGCCATTTGGGCCAGATGGGAGAC 4620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  4778 GTAACCCAGCAGGACGGGGAAGGACCTGGTTTCTACAGCCATTTGGGCCAGATGGGAGAC 4837

Query:  4621 TAAAGACCTGGGTTTACGGTGTAGCAGCTGGGGCATTTGTGTTACTCATCTTTATTGTCT 4680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  4838 TAAAGACCTGGGTTTACGGTGTAGCAGCTGGGGCATTTGTGTTACTCATCTTTATTGTCT 4897

Query:  4681 CCATGATTTATCTAGCTTGCAAAAAGCCAAAGAAACCCCAAAGAAGGCAAAACAACCGAC 4740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  4898 CCATGATTTATCTAGCTTGCAAAAAGCCAAAGAAACCCCAAAGAAGGCAAAACAACCGAC 4957

Query:  4741 TGAAACCTTTAACCTTAGCCTATGATGGAGATGCCGACATGTAACATATAACTTTTCCTG 4800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  4958 TGAAACCTTTAACCTTAGCCTATGATGGAGATGCCGACATGTAACATATAACTTTTCCTG 5017

Query:  4801 GCAACAACCA 4810
             ||||||||||
Sbjct:  5018 GCAACAACCA 5027
```

Table 3. BLASTN VERSUS GENBANK COMPOSITE

>gb:GENBANK-ID:AB023177|acc:AB023177.1 Homo sapiens mRNA for KIAA0960 protein, partial cds - Homo sapiens, 5032 bp. (SEQ ID NO:59)
Length = 5032

Plus Strand HSPs:

Score = 19495 (2925.0 bits), Expect = 0.0, P = 0.0
Identities = 3899/3899 (100%), Positives = 3899/3899 (100%), Strand = Plus / Plus

```
Query:   912 GAGTGGAGCCCCTGCTCAAAAACATGCCATGACATGGTGTCCCCTGCAGGCACTCGTGTA 971
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:     1 GAGTGGAGCCCCTGCTCAAAAACATGCCATGACATGGTGTCCCCTGCAGGCACTCGTGTA 60

Query:   972 AGGACACGAACCATCAGGCAGTTTCCCATTGGCAGTGAAAAGGAGTGTCCAGAATTTGAA 1031
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    61 AGGACACGAACCATCAGGCAGTTTCCCATTGGCAGTGAAAAGGAGTGTCCAGAATTTGAA 120

Query:  1032 GAAAAAGAACCCTGTTTGTCTCAAGGAGATGGAGTTGTCCCCTGTGCCACGTATGGCTGG 1091
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   121 GAAAAAGAACCCTGTTTGTCTCAAGGAGATGGAGTTGTCCCCTGTGCCACGTATGGCTGG 180

Query:  1092 AGAACTACAGAGTGGACTGAGTGCCGTGTGGACCCTTTGCTCAGTCAGCAGGACAAGAGG 1151
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   181 AGAACTACAGAGTGGACTGAGTGCCGTGTGGACCCTTTGCTCAGTCAGCAGGACAAGAGG 240

Query:  1152 CGCGGCAACCAGACGGCCCTCTGTGGAGGGGGCATCCAGACCCGAGAGGTGTACTGCGTG 1211
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   241 CGCGGCAACCAGACGGCCCTCTGTGGAGGGGGCATCCAGACCCGAGAGGTGTACTGCGTG 300

Query:  1212 CAGGCCAACGAAAACCTCCTCTCACAATTAAGTACCCACAAGAACAAAGAAGCCTCAAAG 1271
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   301 CAGGCCAACGAAAACCTCCTCTCACAATTAAGTACCCACAAGAACAAAGAAGCCTCAAAG 360

Query:  1272 CCAATGGACTTAAAATTATGCACTGGACCTATCCCTAATACTACACAGCTGTGCCACATT 1331
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   361 CCAATGGACTTAAAATTATGCACTGGACCTATCCCTAATACTACACAGCTGTGCCACATT 420

Query:  1332 CCTTGTCCAACTGAATGTGAAGTTTCACCTTGGTCAGCTTGGGGACCTTGTACTTATGAA 1391
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   421 CCTTGTCCAACTGAATGTGAAGTTTCACCTTGGTCAGCTTGGGGACCTTGTACTTATGAA 480

Query:  1392 AACTGTAATGATCAGCAAGGGAAAAAAGGCTTCAAACTGAGGAAGCGGCGCATTACCAAT 1451
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   481 AACTGTAATGATCAGCAAGGGAAAAAAGGCTTCAAACTGAGGAAGCGGCGCATTACCAAT 540

Query:  1452 GAGCCCACTGGAGGCTCTGGGGTAACCGGAAACTGCCCTCACTTACTGGAAGCCATTCCC 1511
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   541 GAGCCCACTGGAGGCTCTGGGGTAACCGGAAACTGCCCTCACTTACTGGAAGCCATTCCC 600
```

```
Query: 1512 TGTGAAGAGCCTGCCTGTTATGACTGGAAAGCGGTGAGACTGGGAGACTGCGAGCCAGAT 1571
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  601 TGTGAAGAGCCTGCCTGTTATGACTGGAAAGCGGTGAGACTGGGAGACTGCGAGCCAGAT 660

Query: 1572 AACGGAAAGGAGTGTGGTCCAGGCACGCAAGTTCAAGAGGTTGTGTGCATCAACAGTGAT 1631
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  661 AACGGAAAGGAGTGTGGTCCAGGCACGCAAGTTCAAGAGGTTGTGTGCATCAACAGTGAT 720

Query: 1632 GGAGAAGAAGTTGACAGACAGCTGTGCAGAGATGCCATCTTCCCCATCCCTGTGGCCTGT 1691
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  721 GGAGAAGAAGTTGACAGACAGCTGTGCAGAGATGCCATCTTCCCCATCCCTGTGGCCTGT 780

Query: 1692 GATGCCCCATGCCCGAAAGACTGTGTGCTCAGCACATGGTCTACGTGGTCCTCCTGCTCA 1751
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  781 GATGCCCCATGCCCGAAAGACTGTGTGCTCAGCACATGGTCTACGTGGTCCTCCTGCTCA 840

Query: 1752 CACACCTGCTCAGGGAAAACGACAGAAGGGAAACAGATACGAGCACGATCCATTCTGGCC 1811
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  841 CACACCTGCTCAGGGAAAACGACAGAAGGGAAACAGATACGAGCACGATCCATTCTGGCC 900

Query: 1812 TATGCGGGTGAAGAAGGTGGAATTCGCTGTCCAAATAGCAGTGCTTTGCAAGAAGTACGA 1871
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  901 TATGCGGGTGAAGAAGGTGGAATTCGCTGTCCAAATAGCAGTGCTTTGCAAGAAGTACGA 960

Query: 1872 AGCTGTAATGAGCATCCTTGCACAGTGTACCACTGGCAAACTGGTCCCTGGGGCCAGTGC 1931
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  961 AGCTGTAATGAGCATCCTTGCACAGTGTACCACTGGCAAACTGGTCCCTGGGGCCAGTGC 1020

Query: 1932 ATTGAGGACACCTCAGTATCGTCCTTCAACACAACTACGACTTGGAATGGGGAGGCCTCC 1991
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1021 ATTGAGGACACCTCAGTATCGTCCTTCAACACAACTACGACTTGGAATGGGGAGGCCTCC 1080

Query: 1992 TGCTCTGTCGGCATGCAGACAAGAAAAGTCATCTGTGTGCGAGTCAATGTGGGCCAAGTG 2051
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1081 TGCTCTGTCGGCATGCAGACAAGAAAAGTCATCTGTGTGCGAGTCAATGTGGGCCAAGTG 1140

Query: 2052 GGACCCAAAAAATGTCCTGAAAGCCTTCGACCTGAAACTGTAAGGCCTTGTCTGCTTCCT 2111
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1141 GGACCCAAAAAATGTCCTGAAAGCCTTCGACCTGAAACTGTAAGGCCTTGTCTGCTTCCT 1200

Query: 2112 TGTAAGAAGGACTGTATTGTGACCCCATATAGTGACTGGACATCATGCCCCTCTTCGTGT 2171
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1201 TGTAAGAAGGACTGTATTGTGACCCCATATAGTGACTGGACATCATGCCCCTCTTCGTGT 1260

Query: 2172 AAAGAAGGGGACTCCAGTATCAGGAAGCAGTCTAGGCATCGGGTCATCATTCAGCTGCCA 2231
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1261 AAAGAAGGGGACTCCAGTATCAGGAAGCAGTCTAGGCATCGGGTCATCATTCAGCTGCCA 1320

Query: 2232 GCCAACGGGGGCCGAGACTGCACAGATCCCCTCTATGAAGAGAAGGCCTGTGAGGCACCT 2291
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1321 GCCAACGGGGGCCGAGACTGCACAGATCCCCTCTATGAAGAGAAGGCCTGTGAGGCACCT 1380

Query: 2292 CAAGCGTGCCAAAGCTACAGGTGGAAGACTCACAAATGGCGCAGATGCCAATTAGTCCCT 2351
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1381 CAAGCGTGCCAAAGCTACAGGTGGAAGACTCACAAATGGCGCAGATGCCAATTAGTCCCT 1440

Query: 2352 TGGAGCGTGCAACAAGACAGCCCTGGAGCACAGGAAGGCTGTGGGCCTGGGCGACAGGCA 2411
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1441 TGGAGCGTGCAACAAGACAGCCCTGGAGCACAGGAAGGCTGTGGGCCTGGGCGACAGGCA 1500

Query: 2412 AGAGCCATTACTTGTCGCAAGCAAGATGGAGGACAGGCTGGAATCCATGAGTGCCTACAG 2471
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1501 AGAGCCATTACTTGTCGCAAGCAAGATGGAGGACAGGCTGGAATCCATGAGTGCCTACAG 1560

Query: 2472 TATGCAGGCCCTGTGCCAGCCCTTACCCAGGCCTGCCAGATCCCCTGCCAGGATGACTGT 2531
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1561 TATGCAGGCCCTGTGCCAGCCCTTACCCAGGCCTGCCAGATCCCCTGCCAGGATGACTGT 1620
```

```
Query: 2532 CAATTGACCAGCTGGTCCAAGTTTTCTTCATGCAATGGAGACTGTGGTGCAGTTAGGACC 2591
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1621 CAATTGACCAGCTGGTCCAAGTTTTCTTCATGCAATGGAGACTGTGGTGCAGTTAGGACC 1680

Query: 2592 AGAAAGCGCACTCTTGTTGGAAAAAGTAAAAAGAAGGAAAAATGTAAAAATTCCCATTTG 2651
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1681 AGAAAGCGCACTCTTGTTGGAAAAAGTAAAAAGAAGGAAAAATGTAAAAATTCCCATTTG 1740

Query: 2652 TATCCCCTGATTGAGACTCAGTATTGTCCTTGTGACAAATATAATGCACAACCTGTGGGG 2711
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1741 TATCCCCTGATTGAGACTCAGTATTGTCCTTGTGACAAATATAATGCACAACCTGTGGGG 1800

Query: 2712 AACTGGTCAGACTGTATTTTACCAGAGGGAAAAGTGGAAGTGTTGCTGGGAATGAAAGTA 2771
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1801 AACTGGTCAGACTGTATTTTACCAGAGGGAAAAGTGGAAGTGTTGCTGGGAATGAAAGTA 1860

Query: 2772 CAAGGAGACATCAAGGAATGCGGACAAGGATATCGTTACCAAGCAATGGCATGCTACGAT 2831
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1861 CAAGGAGACATCAAGGAATGCGGACAAGGATATCGTTACCAAGCAATGGCATGCTACGAT 1920

Query: 2832 CAAAATGGCAGGCTTGTGGAAACATCTAGATGTAACAGCCATGGTTACATTGAGGAGGCC 2891
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1921 CAAAATGGCAGGCTTGTGGAAACATCTAGATGTAACAGCCATGGTTACATTGAGGAGGCC 1980

Query: 2892 TGCATCATCCCCTGCCCCTCAGACTGCAAGCTCAGTGAGTGGTCCAACTGGTCGCGCTGC 2951
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1981 TGCATCATCCCCTGCCCCTCAGACTGCAAGCTCAGTGAGTGGTCCAACTGGTCGCGCTGC 2040

Query: 2952 AGCAAGTCCTGTGGGAGTGGTGTGAAGGTTCGTTCTAAATGGCTGCGTGAAAAACCATAT 3011
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2041 AGCAAGTCCTGTGGGAGTGGTGTGAAGGTTCGTTCTAAATGGCTGCGTGAAAAACCATAT 2100

Query: 3012 AATGGAGGAAGGCCTTGCCCCAAACTGGACCATGTCAACCAGGCACAGGTGTATGAGGTT 3071
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2101 AATGGAGGAAGGCCTTGCCCCAAACTGGACCATGTCAACCAGGCACAGGTGTATGAGGTT 2160

Query: 3072 GTCCCATGCCACAGTGACTGCAACCAGTACCTATGGGTCACAGAGCCCTGGAGCATCTGC 3131
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2161 GTCCCATGCCACAGTGACTGCAACCAGTACCTATGGGTCACAGAGCCCTGGAGCATCTGC 2220

Query: 3132 AAGGTGACCTTTGTGAATATGCGGGAGAACTGTGGAGAGGGCGTGCAAACCCGAAAAGTG 3191
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2221 AAGGTGACCTTTGTGAATATGCGGGAGAACTGTGGAGAGGGCGTGCAAACCCGAAAAGTG 2280

Query: 3192 AGATGCATGCAGAATACAGCAGATGGCCCTTCTGAACATGTAGAGGATTACCTCTGTGAC 3251
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2281 AGATGCATGCAGAATACAGCAGATGGCCCTTCTGAACATGTAGAGGATTACCTCTGTGAC 2340

Query: 3252 CCAGAAGAGATGCCCCTGGGCTCTAGAGTGTGCAAATTACCATGCCCTGAGGACTGTGTG 3311
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2341 CCAGAAGAGATGCCCCTGGGCTCTAGAGTGTGCAAATTACCATGCCCTGAGGACTGTGTG 2400

Query: 3312 ATATCTGAATGGGGTCCATGGACCCAATGTGTTTTGCCTTGCAATCAAAGCAGTTTCCGG 3371
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2401 ATATCTGAATGGGGTCCATGGACCCAATGTGTTTTGCCTTGCAATCAAAGCAGTTTCCGG 2460

Query: 3372 CAAAGGTCAGCTGATCCCATCAGACAACCAGCTGATGAAGGAAGATCTTGCCCTAATGCT 3431
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2461 CAAAGGTCAGCTGATCCCATCAGACAACCAGCTGATGAAGGAAGATCTTGCCCTAATGCT 2520

Query: 3432 GTTGAGAAAGAACCCTGTAACCTGAACAAAAACTGCTACCACTATGATTATAATGTAACA 3491
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2521 GTTGAGAAAGAACCCTGTAACCTGAACAAAAACTGCTACCACTATGATTATAATGTAACA 2580

Query: 3492 GACTGGAGTACATGTCAGCTGAGTGAGAAGGCAGTTTGTGGAAATGGAATAAAAACAAGG 3551
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 2581 GACTGGAGTACATGTCAGCTGAGTGAGAAGGCAGTTTGTGGAAATGGAATAAAAACAAGG 2640
```

```
Query:  3552 ATGTTGGATTGTGTTCGAAGTGATGGCAAGTCAGTTGACCTGAAATATTGTGAAGCGCTT 3611
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  2641 ATGTTGGATTGTGTTCGAAGTGATGGCAAGTCAGTTGACCTGAAATATTGTGAAGCGCTT 2700

Query:  3612 GGCTTGGAGAAGAACTGGCAGATGAACACGTCCTGCATGGTGGAATGCCCTGTGAACTGT 3671
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  2701 GGCTTGGAGAAGAACTGGCAGATGAACACGTCCTGCATGGTGGAATGCCCTGTGAACTGT 2760

Query:  3672 CAGCTTTCTGATTGGTCTCCTTGGTCAGAATGTTCTCAAACATGTGGCCTCACAGGAAAA 3731
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  2761 CAGCTTTCTGATTGGTCTCCTTGGTCAGAATGTTCTCAAACATGTGGCCTCACAGGAAAA 2820

Query:  3732 ATGATCCGAAGACGAACAGTGACCCAGCCCTTTCAAGGTGATGGAAGACCATGCCCTTCC 3791
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  2821 ATGATCCGAAGACGAACAGTGACCCAGCCCTTTCAAGGTGATGGAAGACCATGCCCTTCC 2880

Query:  3792 CTGATGGACCAGTCCAAACCCTGCCCAGTGAAGCCTTGTTATCGGTGGCAATATGGCCAG 3851
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  2881 CTGATGGACCAGTCCAAACCCTGCCCAGTGAAGCCTTGTTATCGGTGGCAATATGGCCAG 2940

Query:  3852 TGGTCTCCATGCCAAGTGCAGGAGGCCCAGTGTGGAGAAGGGACCAGAACAAGGAACATT 3911
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  2941 TGGTCTCCATGCCAAGTGCAGGAGGCCCAGTGTGGAGAAGGGACCAGAACAAGGAACATT 3000

Query:  3912 TCTTGTGTAGTAAGTGATGGGTCAGCTGATGATTTCAGCAAAGTGGTGGATGAGGAATTC 3971
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  3001 TCTTGTGTAGTAAGTGATGGGTCAGCTGATGATTTCAGCAAAGTGGTGGATGAGGAATTC 3060

Query:  3972 TGTGCTGACATTGAACTCATTATAGATGGTAATAAAAATATGGTTCTGGAGGAATCCTGC 4031
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  3061 TGTGCTGACATTGAACTCATTATAGATGGTAATAAAAATATGGTTCTGGAGGAATCCTGC 3120

Query:  4032 AGCCAGCCTTGCCCAGGTGACTGTTATTTGAAGGACTGGTCTTCCTGGAGCCTGTGTCAG 4091
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  3121 AGCCAGCCTTGCCCAGGTGACTGTTATTTGAAGGACTGGTCTTCCTGGAGCCTGTGTCAG 3180

Query:  4092 CTGACCTGTGTGAATGGTGAGGATCTAGGCTTTGGTGGAATACAGGTCAGATCCAGACCG 4151
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  3181 CTGACCTGTGTGAATGGTGAGGATCTAGGCTTTGGTGGAATACAGGTCAGATCCAGACCG 3240

Query:  4152 GTGATTATACAAGAACTAGAGAATCAGCATCTGTGCCCAGAGCAGATGTTAGAAACAAAA 4211
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  3241 GTGATTATACAAGAACTAGAGAATCAGCATCTGTGCCCAGAGCAGATGTTAGAAACAAAA 3300

Query:  4212 TCATGTTATGATGGACAGTGCTATGAATATAAATGGATGGCCAGTGCTTGGAAGGGCTCT 4271
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  3301 TCATGTTATGATGGACAGTGCTATGAATATAAATGGATGGCCAGTGCTTGGAAGGGCTCT 3360

Query:  4272 TCCCGAACAGTGTGGTGTCAAAGGTCAGATGGTATAAATGTAACAGGGGGCTGCTTGGTG 4331
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  3361 TCCCGAACAGTGTGGTGTCAAAGGTCAGATGGTATAAATGTAACAGGGGGCTGCTTGGTG 3420

Query:  4332 ATGAGCCAGCCTGATGCCGACAGGTCTTGTAACCCACCGTGTAGTCAACCCCACTCGTAC 4391
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  3421 ATGAGCCAGCCTGATGCCGACAGGTCTTGTAACCCACCGTGTAGTCAACCCCACTCGTAC 3480

Query:  4392 TGTAGCGAGACAAAAACATGCCATTGTGAAGAAGGGTACACTGAAGTCATGTCTTCTAAC 4451
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  3481 TGTAGCGAGACAAAAACATGCCATTGTGAAGAAGGGTACACTGAAGTCATGTCTTCTAAC 3540

Query:  4452 AGCACCCTTGAGCAATGCACACTTATCCCCGTGGTGGTATTACCCACCATGGAGGACAAA 4511
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  3541 AGCACCCTTGAGCAATGCACACTTATCCCCGTGGTGGTATTACCCACCATGGAGGACAAA 3600

Query:  4512 AGAGGAGATGTGAAAACCAGTCGGGCTGTACATCCAACCCAACCCTCCAGTAACCCAGCA 4571
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  3601 AGAGGAGATGTGAAAACCAGTCGGGCTGTACATCCAACCCAACCCTCCAGTAACCCAGCA 3660
```

```
Query:  4572 GGACGGGGAAGGACCTGGTTTCTACAGCCATTTGGGCCAGATGGGAGACTAAAGACCTGG 4631
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  3661 GGACGGGGAAGGACCTGGTTTCTACAGCCATTTGGGCCAGATGGGAGACTAAAGACCTGG 3720

Query:  4632 GTTTACGGTGTAGCAGCTGGGGCATTTGTGTTACTCATCTTTATTGTCTCCATGATTTAT 4691
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  3721 GTTTACGGTGTAGCAGCTGGGGCATTTGTGTTACTCATCTTTATTGTCTCCATGATTTAT 3780

Query:  4692 CTAGCTTGCAAAAAGCCAAAGAAACCCCAAAGAAGGCAAAACAACCGACTGAAACCTTTA 4751
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  3781 CTAGCTTGCAAAAAGCCAAAGAAACCCCAAAGAAGGCAAAACAACCGACTGAAACCTTTA 3840

Query:  4752 ACCTTAGCCTATGATGGAGATGCCGACATGTAACATATAACTTTTCCTGGCAACAACCA 4810
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  3841 ACCTTAGCCTATGATGGAGATGCCGACATGTAACATATAACTTTTCCTGGCAACAACCA 3899
```

SECP11

A SECP11 nucleic acid and polypeptide according to the invention includes the nucleic acid sequence (SEQ ID NO:42 and encoded polypeptide sequence (SEQ ID NO:43) of clone CG50817-04 directed toward novel peptidase (HPEP-8)-like proteins and nucleic acids encoding them. FIG. 16 illustrates the nucleic acid sequence and amino acid sequences. This clone includes a nucleotide sequence (SEQ ID NO:42) of 1447 bp. The nucleotide sequence includes an open reading frame (ORF) beginning with an ATG initiation codon encoding a polypeptide of 224 amino acid residues (SEQ ID NO:43). The start codon is located at nucleotides 520-522 and the stop codon is located at nucleotides 1192-1194. Putative untranslated regions, if any, are found upstream from the initiation codon and downstream from the termination codon. The protein encoded by clone CG50817-04 is predicted by the PSORT program to localize in the cytoplasm with a certainty of 0.4500. The program PSORT and program SignalP predict that the protein appears to have no amino-terminal signal sequence.

Novel peptidase (HPEP-8)-like proteins are related to conditions of failure to thrive, nutritional edema, and hypoproteinemia with normal sweat electrolytes as reported by Townes et al (J. Pediat. 71: 220-224, 1967) for 2 affected male infants. This condition could be treated by a protein hydrolysate diet. Morris and Fisher (Am. J. Dis. Child. 114: 203-208, 1967) reported an affected female who also had imperforate anus, a result of a defect in the synthesis of the enterokinase which activates proteolytic enzymes produced by the pancreas. Oral pancreatin represents a therapeutically successful form of enzyme replacement. Trypsin, like elastase is a member of the pancreatic family of serine proteases. MacDonald et al. (J. Biol. Chem. 257: 9724-9732, 1982) reported nucleotide sequences of cDNAs representing 2 pancreatic rat trypsinogens. The trypsin gene is on mouse chromosome 6 (Honey et al., Somat. Cell Molec.

Genet. 10: 369-376, 1984). Carboxypeptidase A and trypsin are a syntenic pair conserved in mouse and man. Emi et al. (Gene 41: 305-310, 1986) isolated cDNA clones for 2 major human trypsinogen isozymes from a pancreatic cDNA library. The deduced amino acid sequences had 89% homology and the same number of amino acids (247), including a 15-amino acid signal peptide and an 8-amino acid activation peptide. Southern blot analysis of human genomic DNA with the cloned cDNA as a probe showed that the human trypsinogen genes constitute a family of more than 10. The gene encoding trypsin-1 (TRY1) is also referred to as serine protease-1 (PRSS1). Rowen et al. (Science 272: 1755-1762, 1996) found that there are 8 trypsinogen genes embedded in the beta T-cell receptor locus or cluster of genes (TCRB) mapping to 7q35. In the 685-kb DNA segment that they sequenced they found 5 tandemly arrayed 10-kb locus-specific repeats (homology units) at the 3-prime end of the locus. These repeats exhibited 90 to 91% overall nucleotide similarity, and embedded within each is a trypsinogen gene. Alignment of pancreatic trypsinogen cDNAs with the germline sequences showed that these trypsinogen genes contain 5 exons that span approximately 3.6 kb. They denoted 8 trypsinogen genes T1 through T8 from 5-prime to 3-prime. Some of the trypsinogen genes are expressed in nonpancreatic tissues where their function is unknown. Rowen et al. (Science 272: 1755-1762, 1996) noted that the intercalation of the trypsinogen genes in the TCRB locus is conserved in mouse and chicken, suggesting shared functional or regulatory constraints, as has been postulated for genes in the major histocompatibility complex (such as class I, II, and III genes) that share similar long-term organizational relationships. The gene of invention is a novel serine protease containing a trypsin domain but localized on chromosome 16.

The sequence of the invention was derived by laboratory cloning of cDNA fragments covering the full length and/or part of the DNA sequence of the invention, and/or by in silico prediction of the full length and/or part of the DNA sequence of the invention from public human sequence databases.

The laboratory cloning was performed using one or more of the methods summarized as: SeqCallingTM Technology, where cDNA was derived from various human samples representing multiple tissue types, normal and diseased states, physiological states, and developmental states from different donors. Samples were obtained as whole tissue, cell lines, primary cells or tissue cultured primary cells and cell lines. Cells and cell lines may have been treated with biological or chemical agents that regulate gene expression for example, growth factors, chemokines, steroids. The cDNA thus derived was then sequenced using CuraGen's proprietary SeqCalling technology. Sequence traces were evaluated manually and edited for corrections if appropriate. cDNA

sequences from all samples were assembled with themselves and with public ESTs using bioinformatics programs to generate CuraGen's human SeqCalling database of SeqCalling assemblies. Each assembly contains one or more overlapping cDNA sequences derived from one or more human samples. Fragments and ESTs were included as components for an assembly when the extent of identity with another component of the assembly was at least 95% over 50 bp. Each assembly can represent a gene and/or its variants such as splice forms and/or single nucleotide polymorphisms (SNPs) and their combinations.

Exon Linking, where the cDNA coding for the sequence was cloned by polymerase chain reaction (PCR) using the following primers: 5'CTGCTGACCAACACAGCTGCTCAC3' (SEQ ID NO:113) and 5'GACAGGGGCAGTAATGCCATTTGC3' (SEQ ID NO:102) on the following pools of human cDNAs: Pool 1 - Adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus.

Primers were designed based on in silico predictions for the full length or part (one or more exons) of the DNA/protein sequence of the invention or by translated homology of the predicted exons to closely related human sequences or to sequences from other species. Usually multiple clones were sequenced to derive the sequence which was then assembled similar to the SeqCalling process. In addition, sequence traces were evaluated manually and edited for corrections if appropriate.

Variant sequences are also included in this application. A variant sequence can include a single nucleotide polymorphism (SNP). A SNP can, in some instances, be referred to as a "cSNP" to denote that the nucleotide sequence containing the SNP originates as a cDNA. A SNP can arise in several ways. For example, a SNP may be due to a substitution of one nucleotide for another at the polymorphic site. Such a substitution can be either a transition or a transversion. A SNP can also arise from a deletion of a nucleotide or an insertion of a nucleotide, relative to a reference allele. In this case, the polymorphic site is a site at which one allele bears a gap with respect to a particular nucleotide in another allele. SNPs occurring within genes may result in an alteration of the amino acid encoded by the gene at the position of the SNP. Intragenic SNPs may also be silent, however, in the case that a codon including a SNP encodes the same amino acid as

a result of the redundancy of the genetic code. SNPs occurring outside the region of a gene, or in an intron within a gene, do not result in changes in any amino acid sequence of a protein but may result in altered regulation of the expression pattern for example, alteration in temporal expression, physiological response regulation, cell type expression regulation, intensity of expression, stability of transcribed message.

The DNA sequence and protein sequence for a novel Peptidase (HPEP-8)-like gene or one of its splice forms was obtained solely by exon linking and is reported here as CuraGen Acc. No. CG50817-04.

Real-time expression analysis was performed on SECP11 (clone CG50817-04). The results demonstrate that RNA homologous to this clone is present in multiple tissue and cell types.

Accordingly, SECP11 nucleic acids according to the invention can be used to identify one or more of these tissue types. The presence of RNA sequences homologous to a SECP11 nucleic acid in a sample indicates that the sample contains one or more of the above-tissue types.

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 1086 of 1087 bases (99%) identical to a `human peptidase, HPEP-8` mRNA (`patn:A37664`. The full amino acid sequence of the protein of the invention was found to have 254 of 255 amino acid residues (99%) identical to, and 254 of 257 amino acid residues (99%) similar to, the 571 amino acid residue ptnr: `patp:Y41704 Human PRO351 protein sequence` from Homo sapiens.

The presence of identifiable domains in the protein disclosed herein was determined by searches using algorithms such as PROSITE, Blocks, Pfam, ProDomain, Prints and then determining the Interpro number by crossing the domain match (or numbers) using the Interpro website. The results indicate that this protein contains the following protein domains (as defined by Interpro) at the indicated positions: domain name trypsin at amino acid positions 15 to 179. This indicates that the sequence of the invention has properties similar to those of other proteins known to contain this/these domain(s) and similar to the properties of these domains.

Chromosomal information:

The Peptidase (HPEP-8) disclosed in this invention maps to chromosome 16. This information was assigned using OMIM, the electronic northern bioinformatic tool implemented by CuraGen Corporation, public ESTs, public literature references and/or genomic clone

homologies. This was executed to derive the chromosomal mapping of the SeqCalling assemblies, Genomic clones, literature references and/or EST sequences that were included in the invention.

Tissue expression

The Peptidase (HPEP-8) disclosed in this invention is expressed in at least the following tissues: Adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, and/or RACE sources.

Cellular Localization and Sorting

The SignalP, Psort and/or Hydropathy profile for the Peptidase (HPEP-8)-like protein are shown in Table 7. The results predict that this sequence has no signal peptide and is likely to be localized in the cytoplasm with a certainty of 0.4500 predicted by PSORT.

The proteins of the invention encoded by clone CG50817-04 include the protein disclosed as being encoded by the ORF described herein, as well as any mature protein arising therefrom as a result of post-translational modifications. Thus, the proteins of the invention encompass both a precursor and any active forms of the clone CG50817-04 protein.

Functional Variants and Homologs

The novel nucleic acid of the invention encoding a Peptidase (HPEP-8)-like protein includes the nucleic acid whose sequence is provided in Figure 16, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base while still encoding a protein that maintains its Peptidase (HPEP-8)-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications

include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to 1% of the residues may be so changed.

The novel protein of the invention includes the Peptidase (HPEP-8)-like protein whose sequence is provided in Figure 16. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Figure 16 while still encoding a protein that maintains its Peptidase (HPEP-8)-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 1% of the bases may be so changed.

Antibodies

The invention further encompasses antibodies and antibody fragments, such as Fab, (Fab)2 or single chain FV constructs, that bind immunospecifically to any of the proteins of the invention. Also encompassed within the invention are peptides and polypeptides comprising sequences having high binding affinity for any of the proteins of the invention, including such peptides and polypeptides that are fused to any carrier partcle (or biologically expressed on the surface of a carrier) such as a bacteriophage particle.

Uses of the Compositions of the Invention

The protein similarity information, expression pattern, and map location for the Peptidase (HPEP-8)-like protein and nucleic acid disclosed herein suggest that this Peptidase (HPEP-8) may have important structural and/or physiological functions characteristic of the Serine protease family. Therefore, the nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications and as a research tool. These include serving as a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed, as well as potential therapeutic applications such as the following: (i) a protein therapeutic, (ii) a small molecule drug target, (iii) an antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) a nucleic acid

useful in gene therapy (gene delivery/gene ablation), and (v) a composition promoting tissue regeneration in vitro and in vivo (vi) biological defense weapon.

The nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications implicated in various diseases and disorders described below and/or other pathologies. For example, the compositions of the present invention will have efficacy for treatment of patients suffering from: cell proliferative disorder; arteriosclerosis; psoriasis; myelofibrosis; cancer; autoimmune disorder; Crohn's disease; inflammatory disorder; AIDS; anaemia; allergy; asthma; atherosclerosis; Grave's disease; multiple sclerosis; scleroderma; infection; diabetes; metabolic disorder; Addison's disease; cystic fibrosis; glycogen storage disease; obesity; nutritional edema, hypoproteinemia and other diseases, disorders and conditions of the like.

These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods.

Table 4. BLASTN identity search for the nucleic acid of the invention versus GenBank.

>patn:A37664 Human peptidase, HPEP-8 coding sequence - Homo sapiens, 1661 bp. (SEQ ID NO:60)

```
          Length = 1661

  Plus Strand HSPs:

 Score = 5426 (814.1 bits), Expect = 5.1e-240, P = 5.1e-240
 Identities = 1086/1087 (99%), Positives = 1086/1087 (99%), Strand = Plus / Plus

Query:     3 GGACACCAGTGATGCTCCTGGGACCCTACGCAATCTGCGCCTGCGTCTCATCAGTCGCCC 62
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:     1 GGACACCAGTGATGCTCCTGGGACCCTACGCAATCTGCGCCTGCGTCTCATCAGTCGCCC 60

Query:    63 CACATGTAACTGTATCTACAACCAGCTGCACCAGCGACACCTGTCCAACCCGGCCCGGCC 122
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    61 CACATGTAACTGTATCTACAACCAGCTGCACCAGCGACACCTGTCCAACCCGGCCCGGCC 120

Query:   123 TGGGATGCTATGTGGGGGCCCCCAGCCTGGGGTGCAGGGCCCCTGTCAGGTCTGATAGGG 182
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   121 TGGGATGCTATGTGGGGGCCCCCAGCCTGGGGTGCAGGGCCCCTGTCAGGTCTGATAGGG 180

Query:   183 AGAAGAGAAGGAGCAGAAGGGGAGGGGCCTAACCCTGGGCTGGGGGTTGGACTCACAGGA 242
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   181 AGAAGAGAAGGAGCAGAAGGGGAGGGGCCTAACCCTGGGCTGGGGGTTGGACTCACAGGA 240

Query:   243 CTGGGGGAAAGAGCTGCAATCAGAGGGTGTCTGCCATAGCTGGGCTCAGGCATCTGTCCT 302
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   241 CTGGGGGAAAGAGCTGCAATCAGAGGGTGTCTGCCATAGCTGGGCTCAGGCATCTGTCCT 300

Query:   303 TGGCTTTGTTGCCTGGCTCCAGGGAGATTCCGGGGGCCCTGTGCTGTGCCTCGAGCCTGA 362
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   301 TGGCTTTGTTGCCTGGCTCCAGGGAGATTCCGGGGGCCCTGTGCTGTGCCTCGAGCCTGA 360
```

```
Query:   363 CGGACACTGGGTTCAGGCTGGCATCATCAGCTTTGCATCAAGCTGTGCCCAGGAGGACGC 422
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   361 CGGACACTGGGTTCAGGCTGGCATCATCAGCTTTGCATCAAGCTGTGCCCAGGAGGACGC 420

Query:   423 TCCTGTGCTGCTGACCAACACAGCTGCTCACAGTTCCTGGCTGCAGGCTCGAGTTCAGGG 482
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   421 TCCTGTGCTGCTGACCAACACAGCTGCTCACAGTTCCTGGCTGCAGGCTCGAGTTCAGGG 480

Query:   483 GGCAGCTTTCCTGGCCCAGAGCCCAGAGACCCCGGAGATGAGTGATGAGGACAGCTGTGT 542
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   481 GGCAGCTTTCCTGGCCCAGAGCCCAGAGACCCCGGAGATGAGTGATGAGGACAGCTGTGT 540

Query:   543 AGCCTGTGGATCCTTGAGGACAGCAGGTCCCCAGGCAGGAGCACCCTCCCCATGGCCCTG 602
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   541 AGCCTGTGGATCCTTGAGGACAGCAGGTCCCCAGGCAGGAGCACCCTCCCCATGGCCCTG 600

Query:   603 GGAGGCCAGGCTGATGCACCAGGGACAGCTGGCCTGTGGCGGAGCCCTGGTGTCAGAGGA 662
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   601 GGAGGCCAGGCTGATGCACCAGGGACAGCTGGCCTGTGGCGGAGCCCTGGTGTCAGAGGA 660

Query:   663 GGCGGTGCTAACTGCTGCCCACTGCTTCATTGGGCGCCAGGCCCCAGAGGAATGGAGCGT 722
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   661 GGCGGTGCTAACTGCTGCCCACTGCTTCATTGGGCGCCAGGCCCCAGAGGAATGGAGCGT 720

Query:   723 AGGGCTGGGGACCAGACCGGAGGAGTGGGGCCTGAAGCAGCTCATCCTGCATGGAGCCTA 782
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   721 AGGGCTGGGGACCAGACCGGAGGAGTGGGGCCTGAAGCAGCTCATCCTGCATGGAGCCTA 780

Query:   783 CACCCACCCTGAGGGGGGCTACGACATGGCCCTCCTGCTGCTGGCCCAGCCTGTGACACT 842
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   781 CACCCACCCTGAGGGGGGCTACGACATGGCCCTCCTGCTGCTGGCCCAGCCTGTGACACT 840

Query:   843 GGGAGCCAGCCTGCGGCCCCTCTGCCTGCCCTATGCTGACCACCACCTGCCTGATGGGGA 902
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   841 GGGAGCCAGCCTGCGGCCCCTCTGCCTGCCCTATGCTGACCACCACCTGCCTGATGGGGA 900

Query:   903 GCGTGGCTGGGTTCTGGGACGGGCCCGCCCAGGAGCAGGCATCAGCTCCCTCCAGACAGT 962
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   901 GCGTGGCTGGGTTCTGGGACGGGCCCGCCCAGGAGCAGGCATCAGCTCCCTCCAGACAGT 960

Query:   963 GCCCGTGACCCTCCTGGGGCCTAGGGCCTGCAGCCGGCTGCATGCAGCTCCTGGGGGTGA 1022
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   961 GCCCGTGACCCTCCTGGGGCCTAGGGCCTGCAGCCGGCTGCATGCAGCTCCTGGGGGTGA 1020

Query:  1023 TGGCAGCCCTATTCTGCCGGGGATGGTGTGTACCAGTGCTGTGGGTGAGCTGCCCAGCTG 1082
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1021 TGGCAGCCCTATTCTGCCGGGGATGGTGTGTACCAGTGCTGTGGGTGAGCTGCCCAGCTG 1080

Query:  1083 TGAGGCC 1089
             ||||| |
Sbjct:  1081 TGAGGGC 1087

 Score = 1931 (289.7 bits), Expect = 3.7e-82, P = 3.7e-82
 Identities = 635/848 (74%), Positives = 635/848 (74%), Strand = Plus / Plus

Query:   600 CTGGGAGGCCAGGCTGATGCAC-CAGGGACAGCTGGCCTGTGGCGGAGC--CCTGGTGTC 656
             |||   | |||| ||| || || | ||||  ||  ||||| |||    |  ||||   |
Sbjct:   818 CTGCTGGCCCAGCCTG-TG-ACACTGGGA--GCCAGCCTGCGGCCCCTCTGCCTGCCCTA 873

Query:   657 AGAGGAGGCGGTGCTAACTGCTGCCCA-C-TG-CTTCATTGGGCGCCAGGCCC-CAGAGG 712
              |  ||       ||   || ||   | | || ||   ||  |     ||||| |  |||
Sbjct:   874 TGCTGACCACCACCTGCCTGATGGGGAGCGTGGCTGGGTTCTGGGACGGGCCCGCCCAGG 933

Query:   713 AATGGAGCGTAGGGCTGGGGACCAGACCGGAGGAGTGGGGCCTGAAGCAGCTCAT--CCT 770
             |   | ||    |  |     |||||| |     |||   |||   |   |      |||
Sbjct:   934 AGCAG-GCATCAG-CTCCCT-CCAGACAGTGCCCGTGAC-CCTCCTGGGGCCTAGGGCCT 989

Query:   771 GCATGGAGCCTACACCCACCCTGAGGGGGGCTACGACATGGCCCTCCTGCTGCTGGCCCA 830
```

```
            |||  |  | || ||  || |  |    |||||  |  |  ||   |||||  | |||| ||
Sbjct:  990 GCA-GCCGGCTGCATGCAGC-TCCTGGGGGTGATGGCA--GCCCTATT-CTGCCGGGGAT 1044

Query:  831 GCCTGTG-ACACTGGGA-GCCAGCCTGCGGCCCCTCTGCCTGC-CCTATGCTGAC-CACC 886
            |  |||| || | | |  |   |   || ||||  |||   |  ||| | |||   ||||
Sbjct: 1045 GG-TGTGTAC-CAGTGCTGTGGGTGAGCTGCCCAGCTGTGAGGGCCTGT-CTGGGGCACC 1101

Query:  887 ACC--TGCCTGATGGGGAGCGTGGCTGGGTTCTGGGACGGGCCCGCCCAGGAGCAGGCAT 944
            ||   ||| ||| || |||  |   |  |||||  || |||||  || |||   | |  ||
Sbjct: 1102 ACTGGTGCATGA-GGTGAGGGGCACATGGTTCCTGGCCGGGCT-GCACAGCTTCGGAGAT 1159

Query:  945 -CA-GCTCCCTCCA-GACAGTGCCCGTGACCCTCCTGGGGCCTAGGGCCTGCAGCCGGCT 1001
             | ||      ||  | ||| ||| | |  | ||   | ||     ||||   |  | ||
Sbjct: 1160 GCTTGCCAAGGCCCCGCCAG-GCCGGCGGTCTTCACCGCGCTCCCTGCCTAT-GAGGACT 1217

Query: 1002 GCATGCAGCTCCTGGGGGTGATGGCAGCCCTA-TTCTGCCGGGGATGGTGTGTACCAGTG 1060
            |  | ||||   | ||  ||  | |||  ||| ||| |||| ||||    | |  | || |
Sbjct: 1218 GGGT-CAGCAGTTTGGACTG--G-CAGGTCTACTTC-GCCGAGGAACCAGAGCCCGAG-G 1271

Query: 1061 CTGTGGGTG-A-GCTGCCCAGCTGTGAG--GCCAACCAACCAGCTGCTGACAGGGGACCT 1116
            ||| |  || | ||||||  ||    |    ||||||||||||||||||||||||||||||
Sbjct: 1272 CTGAGCCTGGAAGCTGCCTGGCCAACATAAGCCAACCAACCAGCTGCTGACAGGGGACCT 1331

Query: 1117 GGCCATTCTCAGGAACAAGAGAATGCAGGCAGGCAAATGGCATTACTGCCCCTGTCCTCC 1176
            |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1332 GGCCATTCTCAGGA-CAAGAGAATGCAGGCAGGCAAATGGCATTACTGCCCCTGTCCTCC 1390

Query: 1177 CCACCCTGTCATGTGTGATTCCAGGCACCAGGGCAGGCCCAGAAGCCCAGCAGCTGTGGG 1236
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1391 CCACCCTGTCATGTGTGATTCCAGGCACCAGGGCAGGCCCAGAAGCCCAGCAGCTGTGGG 1450

Query: 1237 AAGGAACCTGCCTGGGGCCACAGGTGCCCACTCCCCACCCTGCAGGACAGGGGTGTCTGT 1296
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1451 AAGGAACCTGCCTGGGGCCACAGGTGCCCACTCCCCACCCTGCAGGACAGGGGTGTCTGT 1510

Query: 1297 GGACACTCCCACACCCAACTCTGCTACCAAGCAGGCGTCTCAGCTTTCCTCCTCCTTTAC 1356
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1511 GGACACTCCCACACCCAACTCTGCTACCAAGCAGGCGTCTCAGCTTTCCTCCTCCTTTAC 1570

Query: 1357 CCTTTCAGATACAATCACGCCAGCCACGTTGTTTTGAAAATTTCTTTTTTTGGGGGGCAG 1416
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1571 CCTTTCAGATACAATCACGCCAGCCACGTTGTTTTGAAAATTTCTTTTTTTGGGGGGCAG 1630

Query: 1417 CAGTTTTCCTTTTTTTAAACTTAAATAAATT 1447
            |||||||||||||||||||||||||||||||
Sbjct: 1631 CAGTTTTCCTTTTTTTAAACTTAAATAAATT 1661
```

Table 5. BLASTP identity search for the protein of the invention versus Non-Redundant Composite and GenSeq for the Peptidase (HPEP-8)-like protein of the invention.

```
>patp:Y41704 Human PRO351 protein sequence - Homo sapiens, 571 aa. (SEQ ID
NO:61)

Length = 571

  Plus Strand HSPs:

 Score = 1372 (483.0 bits), Expect = 1.5e-170, Sum P(2) = 1.5e-170
 Identities = 254/255 (99%), Positives = 254/255 (99%), Frame = +1

Query:  322 QGDSGGPVLCLEPDGHWVQAGIISFASSCAQEDAPVLLTNTAAHSSWLQARVQGAAFLAQ 501
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  239 QGDSGGPVLCLEPDGHWVQAGIISFASSCAQEDAPVLLTNTAAHSSWLQARVQGAAFLAQ 298
```

```
Query:   502 SPETPEMSDEDSCVACGSLRTAGPQAGAPSPWPWEARLMHQGQLACGGALVSEEAVLTAA 681
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   299 SPETPEMSDEDSCVACGSLRTAGPQAGAPSPWPWEARLMHQGQLACGGALVSEEAVLTAA 358

Query:   682 HCFIGRQAPEEWSVGLGTRPEEWGLKQLILHGAYTHPEGGYDMALLLAQPVTLGASLRP 861
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   359 HCFIGRQAPEEWSVGLGTRPEEWGLKQLILHGAYTHPEGGYDMALLLAQPVTLGASLRP 418

Query:   862 LCLPYADHHLPDGERGWVLGRARPGAGISSLQTVPVTLLGPRACSRLHAAPGGDGSPILP 1041
             ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   419 LCLPYPDHHLPDGERGWVLGRARPGAGISSLQTVPVTLLGPRACSRLHAAPGGDGSPILP 478

Query:  1042 GMVCTSAVGELPSCE 1086
             |||||||||||||||
Sbjct:   479 GMVCTSAVGELPSCE 493

 Score = 315 (110.9 bits), Expect = 1.5e-170, Sum P(2) = 1.5e-170
 Identities = 56/56 (100%), Positives = 56/56 (100%), Frame = +1

Query:     4 DTSDAPGTLRNLRLRLISRPTCNCIYNQLHQRHLSNPARPGMLCGGPQPGVQGPCQ 171
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   184 DTSDAPGTLRNLRLRLISRPTCNCIYNQLHQRHLSNPARPGMLCGGPQPGVQGPCQ 239

 Score = 225 (79.2 bits), Expect = 8.7e-15, P = 8.7e-15
 Identities = 71/203 (34%), Positives = 95/203 (46%), Frame = +1

Query:   586 PSPWPWEARLMHQGQLACGGALVSEEAVLTAAHCFIGRQAPE--EWSVGLGT------RP 741
             |  |||+| +  ||    | |+||++  ||||||||    | |   ||| ||+       |
Sbjct:    63 PGEWPWQASVRRQGAHICSGSLVADTWVLTAAHCFEKAAATELNSWSVVLGSLQREGLSP 122

Query:   742 --EEWGLKQLILHGAYTHPEGGYDMALLLLAQPVTLGASLRPLCLPYADHHLPDGERGWV 915
               || |+  | |  || |    | |+||| || | |       |||||   |  | | |
Sbjct:   123 GAEEVGVAALQLPRAYNHYSQGSDLALLQLAHPTTH----TPLCLPQPAHRFPFGASCWA 178

Query:   916 LGRARPGAGI-SSLQTVPVTLLGPRACS----RLHAAPGGDGSPILPGMVCTSAVGELPS 1080
              |  +  +    +|+ + + |+    |+    +||     +  |  |||+|    |  |
Sbjct:   179 TGWDQDTSDAPGTLRNLRLRLISRPTCNCIYNQLHQRHLSN--PARPGMLCG---GPQPG 233

Query:  1081 CEANQPAADRGPGHSQEQENAGRQMALLPLSS 1176
              +         ||     | +     |  ++  +|
Sbjct:   234 VQGPCQGDSGGPVLCLEPDGHWVQAGIISFAS 265

 Score = 102 (35.9 bits), Expect = 7.2e-32, Sum P(2) = 7.2e-32
 Identities = 27/84 (32%), Positives = 42/84 (50%), Frame = +1

Query:   295 SVLGFVAWLQGDSGGPVLCLEPDGHWVQAGIISFASSCAQEDAPVLLTNTAAHSSWLQAR 474
             | +| +   +| || | |  |  | |  ||+ ||  +|      | + |   |+  |+ +
Sbjct:   484 SAVGELPSCEGLSGAP-LVHEVRGTWFLAGLHSFGDACQGPARPAVFTALPAYEDWVSS- 541

Query:   475 VQGAAFLAQSPETPEMSDEDSCVA 546
             +    + |+ || || ++  ||+|
Sbjct:   542 LDWQVYFAEEPE-PE-AEPGSCLA 563
```

Table 6. BLASTN identity search (versus the human SeqCalling database for the Peptidase (HPEP-8)-like protein of the invention.

```
>s3aq:132854740 Category D: 12 frag (12 non-5'sig-CG), 636 bp. (SEQ ID NO:62)
           Length = 636

  Minus Strand HSPs:

 Score = 1423 (213.5 bits), Expect = 7.0e-59, P = 7.0e-59
 Identities = 313/343 (91%), Positives = 313/343 (91%), Strand = Minus / Plus
Query:  1001 AGCCGGCTGCAG-GCCCTAGGCCCCAGGAGGGTCACGGGCACTGTCTGGAGGGAGCTGAT 943
             ||| ||||||   | ||| |      ||| || ||   || ||| |    | |  |   |
Sbjct:   295 AGCTGGCTGCCCCGGCCT-GCAGGTTGGATGGACAGCAGCCCTGGCCCT-GTGCCCACCT 352

Query:   942 GCCTGCTCCTGGGCGGGCCCGTCCCAGAACCCAGCCACGCTCCCCATCAGGCAGGTGGTG 883
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Sbjct:  353 ACCTGCTCCTGGGCGGGCCCGTCCCAGAACCCAGCCACGCTCCCCATCAGGCAGGTGGTG 412

Query:  882 GTCAGCATAGGGCAGGCAGAGGGGCCGCAGGCTGGCTCCCAGTGTCACAGGCTGGGCCAG 823
            ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  413 GTCAGGATAGGGCAGGCAGAGGGGCCGCAGGCTGGCTCCCAGTGTCACAGGCTGGGCCAG 472

Query:  822 CAGCAGGAGGGCCATGTCGTAGCCCCCCTCAGGGTGGGTGTAGGCTCCATGCAGGATGAG 763
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  473 CAGCAGGAGGGCCATGTCGTAGCCCCCCTCAGGGTGGGTGTAGGCTCCATGCAGGATGAG 532

Query:  762 CTGCTTCAGGCCCCACTCCTCCGGTCTGGTCCCCAGCCCTACGCTCCATTCCTCTGGGGC 703
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  533 CTGCTTCAGGCCCCACTCCTCCGGTCTGGTCCCCAGCCCTACGCTCCATTCCTCTGGGGC 592

Query:  702 CTGGCGCCCAATGAAGCAGTGGGCAGCAGTTAGCACCGCCTCCT 659
            ||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  593 CTGGCGCCCAATGAAGCAGTGGGCAGCAGTTAGCACCGCCTCCT 636
```

Score = 757 (113.6 bits), Expect = 1.7e-28, P = 1.7e-28 **(SEQ ID NO:103)**
Identities = 165/179 (92%), Positives = 165/179 (92%), Strand = Minus / Plus

```
Query: 1116 AGGTCCCCTGTCAGCAGCTGGTTGGTTGGCCTCACAGCTGGGCAGCTCACCCACAGCACT 1057
            ||||     |||  |     ||||  || |  ||||||||||||||||||||||||||||
Sbjct:  105 AGGTAAGGTGTGGGGGCCTGG--GGCTCACCTCACAGCTGGGCAGCTCACCCACAGCACT 162

Query: 1056 GGTACACACCATCCCCGGCAGAATAGGGCTGCCATCACCCCCAGGAGCTGCATGCAGCCG 997
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  163 GGTACACACCATCCCCGGCAGAATAGGGCTGCCATCACCCCCAGGAGCTGCATGCAGCCG 222

Query:  996 GCTGCAGGCCCTAGGCCCCAGGAGGGTCACGGGCACTGTCTGGAGGGAGCTGATGCCTG 938
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  223 GCTGCAGGCCCTAGGCCCCAGGAGGGTCACGGGCACTGTCTGGAGGGAGCTGATGCCTG 281
```

>s3aq:134913963 Category E: 1 frag (1 non-CG EST), 415 bp.
Length = 415 **(SEQ ID NO:104)**

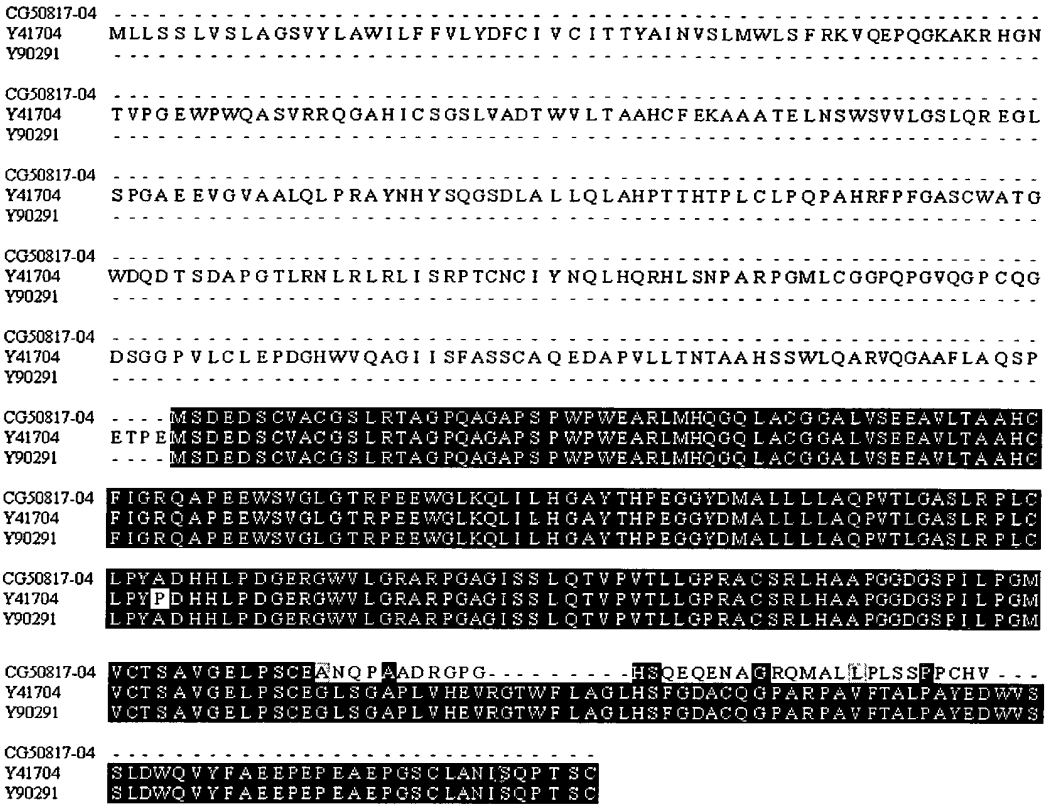
Plus Strand HSPs:

Score = 297 (44.6 bits), Expect = 1.1e-06, P = 1.1e-06
Identities = 61/63 (96%), Positives = 61/63 (96%), Strand = Plus / Plus

```
Query: 1385 TTGTTTTGAAAATTTCTTTTTTTGGGGGGCAGCAGTTTTCCTTTTTTTAAACTTAAATAA 1444
            ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   10 TTGGTGTGAAAATTTCTTTTTTTGGGGGGCAGCAGTTTTCCTTTTTTTAAACTTAAATAA 69

Query: 1445 ATT 1447
            |||
Sbjct:   70 ATT 72
```

Table 7. ClustalW alignment of the protein of the invention with similar peptidase (HPEP-8)s.

ClustalW alignment of the protein of the invention.



Information for the ClustalW proteins:

Accno	Common Name	Length
CG50817-04 (SEQ ID NO:43)	novel Peptidase (HPEP-8)-like protein	
Y41704 (SEQ ID NO:122)	Human PRO351 protein sequence.	571
Y90291 (SEQ ID NO:123)	Human peptidase, HPEP-8 protein sequence.	267

In the alignment shown above, black outlined amino acid residues indicate regions of conserved sequence (i.e., regions that may be required to preserve structural or functional properties); greyed amino acid residues can be mutated to a residue with comparable steric and/or chemical properties without altering protein structure or function (e.g. L to V, I, or M);

non-highlighted amino acid residues can potentially be mutated to a much broader extent without altering structure or function. Psort, SignalP and hydropathy results for the Peptidase (HPEP-8)-like protein of the invention.

Table 8. Psort, Signal P and Pfam Results for CG50817-04, Peptidase (HPEP-8)-like Protein.

PSORT data:

cytoplasm --- Certainty=0.4500(Affirmative) < succ>
microbody (peroxisome) --- Certainty=0.3000(Affirmative) < succ>
lysosome (lumen) --- Certainty=0.2415(Affirmative) < succ>
mitochondrial matrix space --- Certainty=0.1000(Affirmative) < succ>

Signal P data:

# Measure	Position	Value	Cutoff	Conclusion
max. C	57	0.130	0.37	NO
max. Y	55	0.066	0.34	NO
max. S	32	0.311	0.88	NO
mean S	1-54	0.142	0.48	NO

PFAM data:

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
trypsin	Trypsin	69.7	2.7e-21	1

SECP12

A SECP12 nucleic acid and polypeptide according to the invention includes the nucleic acid sequence (SEQ ID NO:44) and encoded polypeptide sequence (SEQ ID NO:45) of clone CG50817-05 directed toward novel peptidase (HPEP-8)-like proteins and nucleic acids encoding them. This is a related variant of SECP11, clone CG50817-04. Figure 17 illustrates the nucleic acid sequence and amino acid sequences respectively. This clone includes a nucleotide sequence (SEQ ID NO:44) of 1592 bp. The nucleotide sequence includes an open reading frame (ORF) beginning with an ATG initiation codon at nucleotides 19-21 and ending with a TGA codon at nucleotides 1582-1584. The encoded protein having 521 amino acid residues is presented using the one-letter code in Figure 17.

The protein encoded by clone CG50817-05 is predicted by the PSORT program to localize in the plasma membrane with a certainty of 0.6850, and appears to be a signal protein (see Table 13 below).

The sequence identified by exon linking was extended in silico using information from at least some of the following sources: SeqCalling assemblies 153687026, 152507187, 153485867, 153485864 and genomic clone gb_AC009088.5 .

The genomic clone was analyzed by Genscan, Grail and/or other programs to identify regions that were putative exons, i.e., putantive coding sequences. The clone was also analyzed by TBLASTN, TFASTN, TFASTA, BLASTX and/or other programs, i.e., hybrid to identify genomic regions translating to proteins with similarity to the original protein or protein family of interest. The following genomic sequence was thus included in the invention: gb_AC009088.5 .

The DNA sequence and protein sequence for a novel Peptidase-like gene or one of its splice forms thus derived is reported here as the invention CG50817-05. Genomic clones having regions with 100% identity to the extended sequence thus obtained were identified by BLASTN searches with the extended sequence against human genomic databases. The genomic clone was selected for further analysis because this identity indicates that these clones contain the genomic locus for these SeqCalling assemblies.

The regions defined by all approaches were then manually integrated and manually corrected for apparent inconsistencies that may have arisen, for example, from miscalled bases in the original fragments used, or from discrepancies between predicted homolgy to a protein of similarity to derive the final sequence of the invention CG50817-05 reported here. When necessary, the process to identify and analyze SeqCalling assemblies, ESTs and genomic clones was reiterated to derive the full length sequence.

Similarities

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 1135 of 1140 bases (99%) identical to a gb:GENBANK-ID: `Z34002 human PRO351 nucleotide sequence` mRNA from Homo (Table 9). The full amino acid sequence of the protein of the invention was found to have 476 of 493 amino acid residues (96%) identical to, and 479 of 493 amino acid residues (97%) similar to, the 571 amino acid residue `patp:Y41704 human PRO351 protein` from Homo sapiens (Table 10).

A multiple sequence alignment is given in Table 12, with the protein of the invention being shown on the first line in a ClustalW analysis comparing the protein of the invention with related protein sequences.

The presence of identifiable domains in the protein disclosed herein was determined by searches using algorithms such as PROSITE, Blocks, Pfam, ProDomain, Prints and then determining the Interpro number by crossing the domain match (or numbers) using the Interpro website. The results indicate that this protein contains the following protein domains (as defined by Interpro) at the indicated positions: domain name trypsin at amino acid positions 61 to 279, and 312 to 476. This indicates that the sequence of the invention has properties similar to those of other proteins known to contain this/these domain(s) and similar to the properties of these domains.

Chromosomal information:

The Peptidase disclosed in this invention maps to chromosome 16. This information was assigned using OMIM, the electronic northern bioinformatic tool implemented by CuraGen Corporation, public ESTs, public literature references and/or genomic clone homologies. This was executed to derive the chromosomal mapping of the SeqCalling assemblies, Genomic clones, literature references and/or EST sequences that were included in the invention.

Tissue expression

The Peptidase disclosed in this invention is expressed in at least the following tissues: Adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, and/or RACE sources.

Cellular Localization and Sorting

The SignalP, Psort and/or Hydropathy profile for the Peptidase-like protein are shown in Table 13. The results predict that this sequence has a signal peptide with a cleavage site between

positions 35 and 36 and is likely to be localized at the plasma membrane with a certainty of 0.6850.

Functional Variants and Homologs

The novel nucleic acid of the invention encoding a Peptidase-like protein includes the nucleic acid whose sequence is provided in Figure 17, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Figure 17, while still encoding a protein that maintains its Peptidase-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 1% of the residues may be so changed.

The novel protein of the invention includes the Peptidase-like protein whose sequence is provided in Figure 17. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Figure 17 while still encoding a protein that maintains its Peptidase-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 4% of the bases may be so changed.

Antibodies

The invention further encompasses antibodies and antibody fragments, such as Fab, (Fab)2 or single chain FV constructs, that bind immunospecifically to any of the proteins of the invention. Also encompassed within the invention are peptides and polypeptides comprising sequences having high binding affinity for any of the proteins of the invention, including such

peptides and polypeptides that are fused to any carrier partcle (or biologically expressed on the surface of a carrier) such as a bacteriophage particle.

Uses of the Compositions of the Invention

The protein similarity information, expression pattern, and map location for the Peptidase-like protein and nucleic acid disclosed herein suggest that this Peptidase may have important structural and/or physiological functions characteristic of the Serine protease family. Therefore, the nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications and as a research tool. These include serving as a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed, as well as potential therapeutic applications such as the following: (i) a protein therapeutic, (ii) a small molecule drug target, (iii) an antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) a nucleic acid useful in gene therapy (gene delivery/gene ablation), and (v) a composition promoting tissue regeneration in vitro and in vivo (vi) biological defense weapon.

The nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications implicated in various diseases and disorders described below and/or other pathologies. For example, the compositions of the present invention will have efficacy for treatment of patients suffering from: cell proliferative disorder; arteriosclerosis; psoriasis; myelofibrosis; cancer; autoimmune disorder; Crohn's disease; inflammatory disorder; AIDS; anaemia; allergy; asthma; atherosclerosis; Grave's disease; multiple sclerosis; scleroderma; infection; diabetes; metabolic disorder; Addison's disease; cystic fibrosis; glycogen storage disease; obesity; nutritional edema, hypoproteinemia and other diseases, disorders and conditions of the like.

These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods.

Table 9. BLASTN identity search for the nucleic acid of the invention.

>patn:Z34002 Human PRO351 nucleotide sequence - Homo sapiens, 2365 bp. (SEQ ID NO:63)

Length = 2365

Plus Strand HSPs:

Score = 5649 (847.6 bits), Expect = 4.3e-288, Sum P(2) = 4.3e-288
Identities = 1135/1140 (99%), Positives = 1135/1140 (99%), Strand = Plus / Plus

```
Query:   340 TCCTGCGTGAGGGACTCAGCCCCTGGGGCCGAAGAGGTGGGGGTGGCTGCCCTGCAGTTG 399
             | | |||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Sbjct:   639 TGCAGCGTGAGGGACTCAGCCC-TGGGGCCGAAGAGGTGGGGGTGGCTGCCCTGCAGTTG 697

Query:   400 CCCAGGGCCTATAACCACTACAGCCAGGGCTCAGACCTGGCCCTGCTGCAGCTCGCCCAC 459
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   698 CCCAGGGCCTATAACCACTACAGCCAGGGCTCAGACCTGGCCCTGCTGCAGCTCGCCCAC 757

Query:   460 CCCACGACCCACACACCCCTCTGCCTGCCCCAGCCCGCCCATCGCTTCCCCTTTGGAGCC 519
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   758 CCCACGACCCACACACCCCTCTGCCTGCCCCAGCCCGCCCATCGCTTCCCCTTTGGAGCC 817

Query:   520 TCCTGCTGGGCCACTGGCTGGGATCAGGACACCAGTGATGCTCCTGGGACCCTACGCAAT 579
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   818 TCCTGCTGGGCCACTGGCTGGGATCAGGACACCAGTGATGCTCCTGGGACCCTACGCAAT 877

Query:   580 CTGCGCCTGCGTCTCATCAGTCGCCCCACATGTAACTGTATCTACAACCAGCTGCACCAG 639
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   878 CTGCGCCTGCGTCTCATCAGTCGCCCCACATGTAACTGTATCTACAACCAGCTGCACCAG 937

Query:   640 CGACACCTGTCCAACCCGGCCCGGCCTGGGATGCTATGTGGGGGCCCCCAGCCTGGGGTG 699
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   938 CGACACCTGTCCAACCCGGCCCGGCCTGGGATGCTATGTGGGGGCCCCCAGCCTGGGGTG 997

Query:   700 CAGGGCCCCTGTCAGGGAGATTCCGGGGGCCCTGTGCTGTGCCTCGAGCCTGACGGACAC 759
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   998 CAGGGCCCCTGTCAGGGAGATTCCGGGGGCCCTGTGCTGTGCCTCGAGCCTGACGGACAC 1057

Query:   760 TGGGTTCAGGCTGGCATCATCAGCTTTGCATCAAGCTGTGCCCAGGAGGACGCTCCTGTG 819
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1058 TGGGTTCAGGCTGGCATCATCAGCTTTGCATCAAGCTGTGCCCAGGAGGACGCTCCTGTG 1117

Query:   820 CTGCTGACCAACACAGCTGCTCACAGTTCCTGGCTGCAGGCTCGAGTTCAGGGGGCAGCT 879
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1118 CTGCTGACCAACACAGCTGCTCACAGTTCCTGGCTGCAGGCTCGAGTTCAGGGGGCAGCT 1177

Query:   880 TTCCTGGCCCAGAGCCCAGAGACCCCGGAGATGAGTGATGAGGACAGCTGTGTAGCCTGT 939
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1178 TTCCTGGCCCAGAGCCCAGAGACCCCGGAGATGAGTGATGAGGACAGCTGTGTAGCCTGT 1237

Query:   940 GGATCCTTGAGGACAGCAGGTCCCCAGGCAGGAGCACCCTCCCCATGGCCCTGGGAGGCC 999
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1238 GGATCCTTGAGGACAGCAGGTCCCCAGGCAGGAGCACCCTCCCCATGGCCCTGGGAGGCC 1297


Query:  1000 AGGCTGATGCACCAGGGACAGCTGGCCTGTGGCGGAGCCCTGGTGTCAGAGGAGGCGGTG 1059
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1298 AGGCTGATGCACCAGGGACAGCTGGCCTGTGGCGGAGCCCTGGTGTCAGAGGAGGCGGTG 1357

Query:  1060 CTAACTGCTGCCCACTGCTTCATTGGGCGCCAGGCCCCAGAGGAATGGAGCGTAGGGCTG 1119
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1358 CTAACTGCTGCCCACTGCTTCATTGGGCGCCAGGCCCCAGAGGAATGGAGCGTAGGGCTG 1417

Query:  1120 GGGACCAGACCGGAGGAGTGGGGCCTGAAGCAGCTCATCCTGCATGGAGCCTACACCCAC 1179
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1418 GGGACCAGACCGGAGGAGTGGGGCCTGAAGCAGCTCATCCTGCATGGAGCCTACACCCAC 1477

Query:  1180 CCTGAGGGGGGCTACGACATGGCCCTCCTGCTGCTGGCCCAGCCTGTGACACTGGGAGCC 1239
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Sbjct:  1478 CCTGAGGGGGGCTACGACATGGCCCTCCTGCTGCTGGCCCAGCCTGTGACACTGGGAGCC 1537

Query:  1240 AGCCTGCGGCCCCTCTGCCTGCCCTATGCTGACCACCACCTGCCTGATGGGGAGCGTGGC 1299
             ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Sbjct:  1538 AGCCTGCGGCCCCTCTGCCTGCCCTATCCTGACCACCACCTGCCTGATGGGGAGCGTGGC 1597

Query:  1300 TGGGTTCTGGGACGGCCCGCCCAGGAGCAGGCATCAGCTCCCTCCAGACAGTGCCCGTG 1359
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1598 TGGGTTCTGGGACGGCCCGCCCAGGAGCAGGCATCAGCTCCCTCCAGACAGTGCCCGTG 1657

Query:  1360 ACCCTCCTGGGGCCTAGGGCCTGCAGCCGGCTGCATGCAGCTCCTGGGGGTGATGGCAGC 1419
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1658 ACCCTCCTGGGGCCTAGGGCCTGCAGCCGGCTGCATGCAGCTCCTGGGGGTGATGGCAGC 1717

Query:  1420 CCTATTCTGCCGGGGATGGTGTGTACCAGTGCTGTGGGTGAGCTGCCCAGCTGTGAGGCC 1479
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Sbjct:  1718 CCTATTCTGCCGGGGATGGTGTGTACCAGTGCTGTGGGTGAGCTGCCCAGCTGTGAGGGC 1777

 Score = 948 (142.2 bits), Expect = 3.0e-74, Sum P(2) = 3.0e-74 (SEQ ID NO:105)
 Identities = 882/1448 (60%), Positives = 882/1448 (60%), Strand = Plus / Plus

Query:   110 TCACCACCTATGCTATCAACGTGAGCCTGATGTGGCTCAGTTT-CCGGAAGGTCCAAGAA 168
             | ||| | | |||| |     ||  | | | |  ||||||  | || |  || |   |
Sbjct:   386 TGACCTCATCTGCTTTGCTT-TGGTCTTCAAGCCGCTCAGCGTGCCTGT-GGACAGCGTG 443

Query:   169 CCCCAGGGCCAACCCAAGCCTCAGGAGGGCAACACAGTCCCTGGCGAGTGGCCCTGGCAG 228
              ||| || ||  ||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   444 GCCCCGGCCCC-CCCAAGCCTCAGGAGGGCAACACAGTCCCTGGCGAGTGGCCCTGGCAG 502

Query:   229 GCCAGTGTGAGGAGGCAAGGAGCCCACATCTGCAGCGGCTCCCTGGTGGCAGACACCTGG 288
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   503 GCCAGTGTGAGGAGGCAAGGAGCCCACATCTGCAGCGGCTCCCTGGTGGCAGACACCTGG 562

Query:   289 GTCCTCACTGCTGCCCACTGCTTTGAAAAGGCAGCAGCAACAGAACTGAATTCCTGCGTG 348
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Sbjct:   563 GTCCTCACTGCTGCCCACTGCTTTGAAAAGGCAGCAGCAACAGAACTGAATTCCTG-GTC 621

Query:   349 AGGGACTCAGCCCCTGGGGCCGAAG-AG-GTGGGGGTGGCTGCCCTGCAGTTGCCCAGG- 405
             || |  ||    | ||||  |   | || ||| |||   | ||||||  |   |||
Sbjct:   622 AGTGG-TC----C-TGGGTTCTCTGCAGCGTGAGGGACTCAGCCCTGGGGCCGAAGAGGT 675

Query:   406 GCCTATAACCACTACAGCCAGG-GCTCAGA-CCTGGCCCTGCTGCAGCTCGC-C-CACCC 461
             |    |  |  |  | || ||  || |||  |||    |  || ||||  |  | ||  |
Sbjct:   676 GGGGGTGGCTGCC-CTGC-AGTTGCCCAGGGCCTATAACCACTACAGCCAGGGCTCAGAC 733

Query:   462 CACGACCCACACACCCCTCTGCCTGCCCCAGCCCGCCCATCGCTTCCCCTTTGGA-GCCT 520
             |  | ||  |   |  ||| |||  |||||  |  |||| | |  ||||| ||   |||
Sbjct:   734 CTGGCCCTGCTG-CAGCTC-GCCCACCCCA--CGACCCA-CACA-CCCCTCTGCCTGCC- 786

Query:   521 CCTGCTGGGCCACTGGCTGGGATCAGGA--CACCAG-TGATGCTC---CTGGGACCCT-A 573
             || ||  | |||  |  |    |  |||  | || | ||   | |   |||||| |   |
Sbjct:   787 CCAGCCCGCCCATCGCTTCCCCTTTGGAGCCTCCTGCTGGGCCACTGGCTGGGATCAGGA 846

Query:   574 CGCAA-TC-TGCGCCTGCGTCTCATCAGTCGCCCCACATGTAACTGTATCTACAACCAGC 631
             | | | |  ||| |||| | | | | |  |||    | ||   |||  ||| || | ||
Sbjct:   847 CACCAGTGATGCTCCTGGGACCC-T-A--CGCAAT-C-TGCGCCTGCGTCT-CATC-AGT 898

Query:   632 TGCACCAGCGACACCTGTC-CAAC--CCGGCCCGGCCTGGGATGCTATGTGGGGGCC--C 686
              || |||     | ||||  | ||  || ||    || | ||  |  |||     ||  |
Sbjct:   899 CGCCCCACATGTAACTGTATCTACAACCAGCTGCACCAGCGACACC-TGTCCAACCCGGC 957

Query:   687 CCAGCCTGGGGTGC-A-G-GGCCCCTGTCAGGGAGAT-TCCGGGGGCCCTGTGCTGTGCC 742
             || ||||||| ||| | | ||   |   ||   |   | | ||| ||||||| | | |
Sbjct:   958 CCGGCCTGGGATGCTATGTGGGGGCCCCCAGCCTGGGGTGCAGGGCCCCTGT-CAGGGA- 1015

Query:   743 TCGAGCCTGACGGACACTGGGTTCAGGCT-G-GCATCATCAG-CTTTGCAT-CAAGCTGT 798
               ||  | |  || | ||| | |  | || | || | | | | |  ||  | || ||||
```

```
Sbjct:  1016 --GATTCCGGGGGCC-CTGTGCTGTGCCTCGAGCCTGA-CGGACACTGGGTTCAGGCTG- 1070

Query:   799 GCC-CAGGAGGAC-GCTCCTGTGCTGCTGACCAACACAGCTGCTCACAGTTC--CTGGCT 854
             || ||  ||        ||  |     |||| || |||  |      | |||| | || |  ||| |
Sbjct:  1071 GCATCATCAGCTTTGCATCAA-GCTG-TGCCCAGGAGGAC-GCTC-CTGTGCTGCTGACC 1126

Query:   855 G-CA--G--GCTCG-AGTTCAGGGG-GCAGCTTTCCTGGCCCAGAGCCCAGAGACCCCGG 907
               ||  |  ||||  |||||  ||   ||||  |   |   |   ||| |   |||      || ||
Sbjct:  1127 AACACAGCTGCTCACAGTTCCTGGCTGCAGGCT--CGAGTTCAGGGGGCAGCTTTCCTGG 1184

Query:   908 AGATGAGTGATGAGGACAGCTGTG-T-AGCC-TGTGGATC-CT-TGAGGACAGCAGGTCC 962
               |||     ||| ||  | | | | ||     || |||    || ||   | | |    | |||
Sbjct:  1185 CCCAGAGCCCAGAG-ACCCCGGAGATGAGTGATGAGGACAGCTGTGTAGCCTGTGGATCC 1243

Query:   963 CC-AGGCAGGAGCACCCTCCCCATGGCCCTGGGAGG-CCAGGCTGATGCACCAGGGACAG 1020
                 |||  |  ||||   |||||| |||     ||||   ||    |    |||   || ||||
Sbjct:  1244 TTGAGG-AC-AGCAGG-TCCCCA-GGCA---GGAGCACCCTCCCCATGGCCCTGGGAGGC 1296

Query:  1021 CTGGCCTGTGGCGG-AGCC-CTGGTGTCAGAGGAGGCGGTGCTAACTGCTGCCCACTGCT 1078
             | ||| ||  ||   ||    | | || |     | ||||| ||    ||| || |       |
Sbjct:  1297 CAGGC-TGATGCACCAGGGACAGCTGGCCT--GTGGCGGAGCC--CTGGTGTCAGAGGAG 1351

Query:  1079 TCATTGGGCGCCAG-GCCC-CAGAGGAATGGAGCGT-AGGGCTG-G-GGACCAGACCGGA 1133
               |  ||     |     |||| | |     || | |||  ||| |     | |||    || || |
Sbjct:  1352 GCGGTGCTAACTGCTGCCCACTGCTTCATTGGGCGCCAGGCCCCAGAGGAATGGAGCGTA 1411

Query:  1134 GGAGTGGGG-CCTGAAGCAGCTCA-TCCTGCATGGAGCCTACACCCACCCTG-AGGGGGG 1190
             ||  ||||| || ||   |  |     | |     || || |||    |    | || |||| | || |
Sbjct:  1412 GGGCTGGGGACCAGAC-CGGAGGAGTGGGGCCTGAAGC--AG-CTCATCCTGCATGGAGC 1467

Query:  1191 CTACGACATGGCCCTCCTGCTG-CTGGCCCA-GCCTGTGACACTGGGAGCC-AGCCTGCG 1247
             |||| ||     ||||    |   | ||     || | |    |      |||    ||| |||||| |
Sbjct:  1468 CTAC-ACCCA-CCCTGAGGGGGGCTACGACATGGCCCTCCTGCTGCTGGCCCAGCCTGTG 1525

Query:  1248 GCCCCTCT-GCCTGCCCTATGCTGACCACCA-CCTGCCTGATGGGGAGCGTGGC-TGGGT 1304
              | |      ||| |||    ||| | ||    |  |||||| ||     || |      | ||   |
Sbjct:  1526 ACACTGGGAGCCAGCC---TGCGGCCCCTCTGCCTGCCCTATCCTGACCACCACCTGCCT 1582

Query:  1305 TCTGGGACGGGCCCGCCCAGGAGCAGGCATCAGCTCCCTCCAGACAGTGCCCGTGACCCT 1364
                 ||||   | ||   |  |   ||   | || |    || |  | ||||   ||       | | | |
Sbjct:  1583 GATGGG--GAGCGTGGCTGGGTTCTGGGACGGGCCCGC-CCAGG-AGCAGGCATCAGCTC 1638

Query:  1365 CCTGGGGCCTAGGGCCTGC-AGCCGGCTGCATGC-AGCTCCTGGGGGTGATGGCAGCCCT 1422
             |||   | | || ||| |   | ||   |||     | ||   ||||   |   |      |||  |
Sbjct:  1639 CCTCCAGAC-AGTGCCCGTGACCCTCCTGGGGCCTAGGGCCTGCAGCCGGCTGCATGCAG 1697

Query:  1423 ATTCTGCCGGGGATGGTGTGTACCAGT--GCTGTGGGTGAGCTGC-CCAG--CTGTGAGG 1477
               | |||   || |||||      | | | |  || | || ||    ||   ||||  ||||| ||
Sbjct:  1698 CTCCTGGGGGTGATGGCA-GCCCTATTCTGCCGGGGATGGTGTGTACCAGTGCTGTG-GG 1755

Query:  1478 CCAACCAACCAGCTGCTGACAGGGGACCTGGC-CATTCTCAGGAACAAGAGAATGCAGGC 1536
               | |  ||||||| |||   |       ||||  ||  | | | ||  || |||  || |||
Sbjct:  1756 TGAGCTGCCCAGCTG-TGAGGGCCTGTCTGGGGCAC-CACTGGTGCATGAGG-TG-AGG- 1810

Query:  1537 AGGCAAATGGCATTACTGCCC 1557
              |||| ||||  || ||| ||
Sbjct:  1811 -GGCACATGG--TTCCTGGCC 1828
```

Score = 894 (134.1 bits), Expect = 4.3e-288, Sum P(2) = 4.3e-288 **(SEQ ID NO:106)**
Identities = 182/186 (97%), Positives = 182/186 (97%), Strand = Plus / Plus

```
Query:     1 CGCTGGGCCTCTGTCCTGATGCTGCTGAGCTCCCTGGTGTCTCTCGCTGGTTCTGTCTAC 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   171 CGCTGGGCCTCTGTCCTGATGCTGCTGAGCTCCCTGGTGTCTCTCGCTGGTTCTGTCTAC 230

Query:    61 CTGGCCTGGATCCTGTTCTTCGTGCTCTATGATTTCTGCATTGTTTGTATCACCACCTAT 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   231 CTGGCCTGGATCCTGTTCTTCGTGCTCTATGATTTCTGCATTGTTTGTATCACCACCTAT 290
```

```
Query:   121 GCTATCAACGTGAGCCTGATGTGGCTCAGTTTCCGGAAGGTCCAAGAACCCCAGGGCCAA 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Sbjct:   291 GCTATCAACGTGAGCCTGATGTGGCTCAGTTTCCGGAAGGTCCAAGAACCCCAGGGCAAG 350

Query:   181 CCCAAG 186
               | |||
Sbjct:   351 GCTAAG 356
```

Score = 699 (104.9 bits), Expect = 9.8e-60, Sum P(2) = 9.8e-60 **(SEQ ID NO:107)**
Identities = 391/603 (64%), Positives = 391/603 (64%), Strand = Plus / Plus

```
Query:   990 CTGGGAGGCCAGGCTGATGCAC-CAGGGACAGCTGGCCTGTGGCGGAGC--CCTGG--TG 1044
             |||   | |||| ||| || || | ||||  ||  ||||| |||    |  ||||   |
Sbjct:  1508 CTGCTGGCCCAGCCTG-TG-ACACTGGGA--GCCAGCCTGCGGCCCCTCTGCCTGCCCTA 1563

Query:  1045 TCA-GAGGAGGCGGTGC-TAACTGCTGCCCACTGCTTCATTGGGCGCCAGGCCC-CAGAG 1101
             ||  ||  |     ||| | | ||  |  |   |||   ||  | | | ||||| |  ||
Sbjct:  1564 TCCTGACCACCACCTGCCTGA-TGGGGAGCGTGGCTGGGTTCTGGGACGGGCCCGCCCAG 1622

Query:  1102 GAATGGAGCGTAGGGCTGGGGACCAGACCGGAGGAGTGGGGCCTGAAGCAGCTCAT--CC 1159
             ||   | || |  | ||      |||||| |      |||   |||   |  ||  |   ||
Sbjct:  1623 GAGCAG-GCATCAG-CTCCCT-CCAGACAGTGCCCGTGAC-CCTCCTGGGGCCTAGGGCC 1678

Query:  1160 TGCATGGAGCCTACACCCACCCTGAGGGGGGCTACGACATGGCCCTCCTGCTGCTGGCCC 1219
             |||| |  | || ||  || | |   |||||  | | ||  |||||  | |||| ||
Sbjct:  1679 TGCA-GCCGGCTGCATGCAGC-TCCTGGGGGTGATGGCA--GCCCTATT-CTGCCGGGGA 1733

Query:  1220 AGCCTGTG-ACACTGGGA-GCCAGCCTGCGGCCCCTCTGCCTGC-CCTATGCTGAC-CAC 1275
              |  |||| || | | |  |   |  || ||||  |||   |  ||| | |||   |||
Sbjct:  1734 TGG-TGTGTAC-CAGTGCTGTGGGTGAGCTGCCCAGCTGTGAGGGCCTGT-CTGGGGCAC 1790

Query:  1276 CACC--TGCCTGATGGGGAGCGTGGCTGGGTTCTGGGACGGGCCCGCCCAGGAGCAGGCA 1333
             |||   ||| ||| || ||| |   |  |||||  || |||||  || |||   | |  |
Sbjct:  1791 CACTGGTGCATGA-GGTGAGGGGCACATGGTTCCTGGCCGGGCT-GCACAGCTTCGGAGA 1848

Query:  1334 T-CA-GCTCCCTCCA-GACAGTGCCCGTGACCCTCCTGGGGCCTAGGGCCTGCAGCCGGC 1390
             | |  ||     ||  | ||| ||| | |  | ||  | ||     ||||   | | |
Sbjct:  1849 TGCTTGCCAAGGCCCCGCCAG-GCCGGCGGTCTTCACCGCGCTCCCTGCCTAT-GAGGAC 1906

Query:  1391 TGCATGCAGCTCCTGGGGGGTGATGGCAGCCCTA-TTCTGCCGGGGATGGTGTGTACCAGT 1449
             ||  | ||||   | ||  ||  | |||  ||| ||| |||| |||    | |  | ||
Sbjct:  1907 TGGGT-CAGCAGTTTGGACTG--G-CAGGTCTACTTC-GCCGAGGAACCAGAGCCCGAG- 1960

Query:  1450 GCTGTGGGTG-A-GCTGCCCAGCTGTGAG--GCCAACCAACCAGCTGCTGACAGGGGACC 1505
             |||| |  || | ||||||  ||    |  ||||||||||||||||||||||||||||
Sbjct:  1961 GCTGAGCCTGGAAGCTGCCTGGCCAACATAAGCCAACCAACCAGCTGCTGACAGGGGACC 2020

Query:  1506 TGGCCATTCTCAGGAACAAGAGAATGCAGGCAGGCAAATGGCATTACTGCCCCTGTCCTC 1565
             |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Sbjct:  2021 TGGCCATTCTCAGGA-CAAGAGAATGCAGGCAGGCAAATGGCATTACTGCCCCTGTCCTC 2079

Query:  1566 CCCACCCTGTCATGTGTGATTCCAGGC 1592
             ||||||||||||||||||||||||||
Sbjct:  2080 CCCACCCTGTCATGTGTGATTCCAGGC 2106
```

>patn:A37664 Human peptidase, HPEP-8 coding sequence - Homo sapiens, 1661 bp.
(SEQ ID NO:64)
Length = 1661

Plus Strand HSPs:

Score = 3831 (574.8 bits), Expect = 5.6e-168, P = 5.6e-168
Identities = 767/768 (99%), Positives = 767/768 (99%), Strand = Plus / Plus

```
Query:   712 CAGGGAGATTCCGGGGGCCCTGTGCTGTGCCTCGAGCCTGACGGACACTGGGTTCAGGCT 771
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   320 CAGGGAGATTCCGGGGGCCCTGTGCTGTGCCTCGAGCCTGACGGACACTGGGTTCAGGCT 379
```

```
Query:   772 GGCATCATCAGCTTTGCATCAAGCTGTGCCCAGGAGGACGCTCCTGTGCTGCTGACCAAC 831
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   380 GGCATCATCAGCTTTGCATCAAGCTGTGCCCAGGAGGACGCTCCTGTGCTGCTGACCAAC 439

Query:   832 ACAGCTGCTCACAGTTCCTGGCTGCAGGCTCGAGTTCAGGGGGCAGCTTTCCTGGCCCAG 891
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   440 ACAGCTGCTCACAGTTCCTGGCTGCAGGCTCGAGTTCAGGGGGCAGCTTTCCTGGCCCAG 499

Query:   892 AGCCCAGAGACCCCGGAGATGAGTGATGAGGACAGCTGTGTAGCCTGTGGATCCTTGAGG 951
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   500 AGCCCAGAGACCCCGGAGATGAGTGATGAGGACAGCTGTGTAGCCTGTGGATCCTTGAGG 559

Query:   952 ACAGCAGGTCCCCAGGCAGGAGCACCCTCCCCATGGCCCTGGGAGGCCAGGCTGATGCAC 1011
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   560 ACAGCAGGTCCCCAGGCAGGAGCACCCTCCCCATGGCCCTGGGAGGCCAGGCTGATGCAC 619

Query:  1012 CAGGGACAGCTGGCCTGTGGCGGAGCCCTGGTGTCAGAGGAGGCGGTGCTAACTGCTGCC 1071
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   620 CAGGGACAGCTGGCCTGTGGCGGAGCCCTGGTGTCAGAGGACGCGGTCCTAACTGCTGCC 679

Query:  1072 CACTGCTTCATTGGGCGCCAGGCCCCAGAGGAATGGAGCGTAGGGCTGGGGACCAGACCG 1131
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   680 CACTGCTTCATTGGGCGCCAGGCCCCAGAGGAATGGAGCGTAGGGCTGGGGACCAGACCG 739

Query:  1132 GAGGAGTGGGGCCTGAAGCAGCTCATCCTGCATGGAGCCTACACCCACCCTGAGGGGGGC 1191
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   740 GAGGAGTGGGGCCTGAAGCAGCTCATCCTGCATGGAGCCTACACCCACCCTGAGGGGGGC 799

Query:  1192 TACGACATGGCCCTCCTGCTGCTGGCCCAGCCTGTGACACTGGGAGCCAGCCTGCGGCCC 1251
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   800 TACGACATGGCCCTCCTGCTGCTGGCCCAGCCTGTGACACTGGGAGCCAGCCTGCGGCCC 859

Query:  1252 CTCTGCCTGCCCTATGCTGACCACCACCTGCCTGATGGGGAGCGTGGCTGGGTTCTGGGA 1311
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   860 CTCTGCCTGCCCTATGCTGACCACCACCTGCCTGATGGGGAGCGTGGCTGGGTTCTGGGA 919

Query:  1312 CGGGCCCGCCCAGGAGCAGGCATCAGCTCCCTCCAGACAGTGCCCGTGACCCTCCTGGGG 1371
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   920 CGGGCCCGCCCAGGAGCAGGCATCAGCTCCCTCCAGACAGTGCCCGTGACCCTCCTGGGG 979

Query:  1372 CCTAGGGCCTGCAGCCGGCTGCATGCAGCTCCTGGGGGTGATGGCAGCCCTATTCTGCCG 1431
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   980 CCTAGGGCCTGCAGCCGGCTGCATGCAGCTCCTGGGGGTGATGGCAGCCCTATTCTGCCG 1039

Query:  1432 GGGATGGTGTGTACCAGTGCTGTGGGTGAGCTGCCCAGCTGTGAGGCC 1479
             |||||||||||||||||||||||||||||||||||||||||||||| |
Sbjct:  1040 GGGATGGTGTGTACCAGTGCTGTGGGTGAGCTGCCCAGCTGTGAGGGC 1087
```

Score = 974 (146.1 bits), Expect = 6.1e-39, P = 6.1e-39 **(SEQ ID NO:108)**
Identities = 632/998 (63%), Positives = 632/998 (63%), Strand = Plus / Plus

```
Query:   546 GGACACCAGTGATGCTCCTGGGACCCTACGCAATCTGCGCCTGCGTCTCATCAGTCGCCC 605
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:     1 GGACACCAGTGATGCTCCTGGGACCCTACGCAATCTGCGCCTGCGTCTCATCAGTCGCCC 60

Query:   606 CACATGTAACTGTATCTACAACCAGCTGCACCAGCGACACCTGTCCAACCCGGCCCGGCC 665
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    61 CACATGTAACTGTATCTACAACCAGCTGCACCAGCGACACCTGTCCAACCCGGCCCGGCC 120

Query:   666 TGGGATGCTATGTGGGGGCCCCCAGCCTGGGGTGCAGGGCCCCTGTCAGGGA-GATTCCG 724
             ||||||||||||||||||||||||||||||||||||||||||||||||||   |||   |
Sbjct:   121 TGGGATGCTATGTGGGGGCCCCCAGCCTGGGGTGCAGGGCCCCTGTCAGGTCTGATAGGG 180

Query:   725 GGG-GCCCTGT-GCTGTGCCTCGAGCCTGACGGACACTGGGTTCAGGCTGGCA-TCATCA 781
              |  |    |  || |      |||   | |   | ||||| |  || | | | ||| ||
Sbjct:   181 AGAAGAGAAGGAGCAGAAGGG-GAGGG-GCCTAACCCTGGGCTGGGGGTTGGACTCA-CA 237
```

```
Query:   782 G--CTTTGCATCA-AGCTGTGCCCAGGAGGACGCTCCTGTGCT-GCTGACCA-ACACAGC 836
Sbjct:   238 GGACTGGGGGAAAGAGCTGCAATCAG-AGGGTG-TC-TGCCATAGCTGGGCTCAGGCATC 294

Query:   837 TGCTCACAGTTCCTGGCTGCA-GGCTC---G-AG-TTCAGGGGGCAGCTTTCCTG-GCCC 889
Sbjct:   295 TG-TCCTTGG-CTTTGTTGCCTGGCTCCAGGGAGATTCCGGGGGCC-CTGTGCTGTGCCT 351

Query:   890 AGAGCCC-AGAGACCCCGGAGATGAGTGATGAGGACAGCTGTGTAGCCTGTGGATCCT-- 946
Sbjct:   352 CGAGCCTGACGGACACTGG-GTTCAG-GCTG--G-CATCA-TC-AGCTT-TGCATCAAGC 403

Query:   947 TGAGGACAGCAGGTC-C-CCAG-GCAGGAGCACCCTCCCCATGGCCCTGGGAGG-CCAGG 1002
Sbjct:   404 TGTGCCCAGGAGGACGCTCCTGTGCTGCTG-ACCAACAC-A-GCTGCTCACAGTTCCTGG 460

Query:  1003 CTG-ATGCACCAGGGACAGCTGGCCTGTGGCGGAGCCCTGGTGTCAGAGGAGGCGGTGCT 1061
Sbjct:   461 CTGCAGGCTCGAGTT-CAGGGGGCAGCTTTCCTGGCCCAGAGCCCAGAGACCCCGGAGAT 519

Query:  1062 AACTGCTGCCCACTGCTTCATTGGGCGCCAGGCCCCAGAGGAATGGAG--CGTAGGGCTG 1119
Sbjct:   520 GAGTGATGAGGACAGCTGTGTAGCCTGTG-GATCCTTGAGGACAGCAGGTCCCCAGGCAG 578

Query:  1120 GGG-ACCAGACCGGAGGAGTGGGGCCTGAAGCAGCTCATCCTGCATGGAGC-CTACACCC 1177
Sbjct:   579 GAGCACCCTCCCCATGGCCCTGGGAGGCCAG--GCTGATGCACCAGGGACAGCTGGCCTG 636

Query:  1178 ACCCTGAGGGGGGCTA-C-GACATGGCCCTCCTG-CTGCTGGCCCAGCCTGTGACACTGG 1234
Sbjct:   637 TGGCGGAGCCCTGGTGTCAGAGGAGGCGGTGCTAACTGCTG-CCCA--CTGCTTCATTGG 693

Query:  1235 GAGCCAGCCTGCGGCCCCTCTGCCTGCCCTATG-CTGACCACCAC-CTGCCTGA-TGGGG 1291
Sbjct:   694 GCGCCAGGCCCCAGAGGAA-TGGA-GCG-TAGGGCTGGGGACCAGACCGGAGGAGTGGGG 750

Query:  1292 AGCGTGGCTGGGT-TCTGGGACGGGCCCGCCCAGGAGCAGGCATCAGCTCC-CTCCAGAC 1349
Sbjct:   751 CCTGAAGCAGCTCATCCTGCATGGAGCCTAC-ACC--CACCC-TGAGGGGGGCTAC-GAC 805

Query:  1350 AGTGCCCGTGACCCTCCTGGG---GCCTAGGGC-CTGC-AGCCGGC-TGCATGCAGCTCC 1403
Sbjct:   806 ATGGCCC-TCCTGCTGCTGGCCCAGCCTGTGACACTGGGAGCCAGCCTGCG-GCCCCTC- 862

Query:  1404 TGGGGGTGATG-GCAG-CC-CTATTCTGCCGGGGATGGTGTGTACCAGTGCTGTGGGT-G 1459
Sbjct:   863 TGCCTGCCCTATGCTGACCACCAC-CTGCCTGATGGGGAGCGTGGCTGGGTTCTGGGACG 921

Query:  1460 AGCT-GCCCAGCTGTGAGGCCAACCAACCAGCTGCTGACAGGGGACCTGGCCATTCTCAG 1518
Sbjct:   922 GGCCCGCCCAGGAGC-AGGC--ATCAGCTCCCTCCAGACAGTGCCCGTGACCCTCCTGGG 978

Query:  1519 GAACAAGAGAATGCAGGCAGGC 1540
Sbjct:   979 GC-CTAGGGCCTGCAGCC-GGC 998
```

Score = 706 (105.9 bits), Expect = 1.9e-23, P = 1.9e-23 **(SEQ ID NO:109)**
Identities = 390/603 (64%), Positives = 390/603 (64%), Strand = Plus / Plus

```
Query:   990 CTGGGAGGCCAGGCTGATGCAC-CAGGGACAGCTGGCCTGTGGCGGAGC--CCTGGTGTC 1046
Sbjct:   818 CTGCTGGCCCAGCCTG-TG-ACACTGGGA--GCCAGCCTGCGGCCCCTCTGCCTGCCCTA 873

Query:  1047 AGAGGAGGCGGTGCTAACTGCTGCCCA-C-TG-CTTCATTGGGCGCCAGGCCC-CAGAGG 1102
Sbjct:   874 TGCTGACCACCACCTGCCTGATGGGGAGCGTGGCTGGGTTCTGGGACGGGCCCGCCCAGG 933

Query:  1103 AATGGAGCGTAGGGCTGGGGACCAGACCGGAGGAGTGGGGCCTGAAGCAGCTCAT--CCT 1160
```

```
             |   | || |  | ||     |||||| |     |||   |||   |  ||  |   |||
Sbjct:   934 AGCAG-GCATCAG-CTCCCT-CCAGACAGTGCCCGTGAC-CCTCCTGGGGCCTAGGGCCT 989

Query:  1161 GCATGGAGCCTACACCCACCCTGAGGGGGGCTACGACATGGCCCTCCTGCTGCTGGCCCA 1220
             ||| |  | || ||  || | |   |||||  | | ||  |||||  | |||| ||
Sbjct:   990 GCA-GCCGGCTGCATGCAGC-TCCTGGGGGTGATGGCA--GCCCTATT-CTGCCGGGGAT 1044

Query:  1221 GCCTGTG-ACACTGGGA-GCCAGCCTGCGGCCCCTCTGCCTGC-CCTATGCTGAC-CACC 1276
             |  |||| || | | |  |   |   || ||||  |||   |  ||| | |||   ||||
Sbjct:  1045 GG-TGTGTAC-CAGTGCTGTGGGTGAGCTGCCCAGCTGTGAGGGCCTGT-CTGGGGCACC 1101

Query:  1277 ACC--TGCCTGATGGGGAGCGTGGCTGGGTTCTGGGACGGGCCCGCCCAGGAGCAGGCAT 1334
             ||   ||| ||| || ||| |   |  |||||  || |||||  || |||   | |  ||
Sbjct:  1102 ACTGGTGCATGA-GGTGAGGGGCACATGGTTCCTGGCCGGGCT-GCACAGCTTCGGAGAT 1159

Query:  1335 -CA-GCTCCCTCCA-GACAGTGCCCGTGACCCTCCTGGGGCCTAGGGCCTGCAGCCGGCT 1391
              |  ||     ||  | ||| ||| | |  | ||   | ||     ||||   |  | ||
Sbjct:  1160 GCTTGCCAAGGCCCCGCCAG-GCCGGCGGTCTTCACCGCGCTCCCTGCCTAT-GAGGACT 1217

Query:  1392 GCATGCAGCTCCTGCCGGTGATGGCAGCCCTA-TTCTGCCGGGGATGGTGTGTACCAGTG 1450
             |  | ||||   |     ||  | |||  ||| ||| |||| |||    | |  | || |
Sbjct:  1218 GGGT-CAGCAGTTTGGACTG--G-CAGGTCTACTTC-GCCGAGGAACCAGAGCCCGAG-G 1271

Query:  1451 CTGTGGGTG-A-GCTGCCCAGCTGTGAG--GCCAACCAACCAGCTGCTGACAGGGGACCT 1506
             ||| |  || | ||||||  ||    |   ||||||||||||||||||||||||||||||
Sbjct:  1272 CTGAGCCTGGAAGCTGCCTGGCCAACATAAGCCAACCAACCAGCTGCTGACAGGGGACCT 1331

Query:  1507 GGCCATTCTCAGGAACAAGAGAATGCAGGCAGGCAAATGGCATTACTGCCCCTGTCCTCC 1566
             |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1332 GGCCATTCTCAGGA-CAAGAGAATGCAGGCAGGCAAATGGCATTACTGCCCCTGTCCTCC 1390

Query:  1567 CCACCCTGTCATGTGTGATTCCAGGC 1592
             ||||||||||||||||||||||||||
Sbjct:  1391 CCACCCTGTCATGTGTGATTCCAGGC 1416
```

Score = 481 (72.2 bits), Expect = 1.1e-12, P = 1.1e-12 **(SEQ ID NO:110)**
Identities = 409/666 (61%), Positives = 409/666 (61%), Strand = Plus / Plus

```
Query:   207 CCCTGGCGAGTGGCCCTGGCAGGCCAGTGTGAGGAGGCAAGGAGCCCACATCTGCAGCGG 266
             ||||  | | ||||||||| |||||||  ||| |   || |||   |    |||  ||||
Sbjct:   584 CCCTCCCCA-TGGCCCTGGGAGGCCAGGCTGATGCACCAGGGACAGCTGGCCTGTGGCGG 642

Query:   267 CTCCCTGGTGGCAGACACCTGGGTCCTCACTGCTGCCCACTGCTTTGAAAAGGCAGCAG- 325
               |||||||| ||||      ||| || |||||||||||||||||  |   |||  |||
Sbjct:   643 AGCCCTGGTGTCAGAGGAGGCGGTGCTAACTGCTGCCCACTGCTTC-ATTGGGCGCCAGG 701

Query:   326 CAACAGAACTGAATTCCTGCGTGAGGGACTCAGCCCCTGGGGCCGAAGAGGTGGGGGTGG 385
             |  ||||   ||||    |||| |||| ||  |  || |   | || ||| |||||   |
Sbjct:   702 CCCCAGAG--GAATGGA-GCGT-AGGG-CTGGGGACCAGAC-CGGAGGAG-TGGGGCCTG 754

Query:   386 CTGCC-CTGCAGT-TGCCCAGGGCCTATAACCACTAC-AGCCAGGGCTCAGACCTGGCCC 442
               ||  || ||   |||   | ||||| | ||||    ||   |||||  ||| ||||||
Sbjct:   755 AAGCAGCT-CATCCTGCATGGAGCCTACACCCACCCTGAGGG-GGGCTACGACATGGCCC 812

Query:   443 TGCTGCAGCTCGCCCACCC-----CAC--G-ACCCA-CA--CA-CCCCTCTGCCTGCCCC 490
             | |||| ||| ||||| ||     |||  | | ||| |   |  |||||||||||||||
Sbjct:   813 TCCTGCTGCTGGCCCAGCCTGTGACACTGGGAGCCAGCCTGCGGCCCCTCTGCCTGCCCT 872

Query:   491 AGCCCGCCCATCGCTTCCCCTTTGGAGCCTCCTG-CTGGGCCACTGGCTGGGATCAGGAC 549
             |  | | ||| | | | ||   ||| |   | || |||||   ||||   ||  | |  |
Sbjct:   873 ATGCTGACCACCACCTGCCTGATGGGGAG-CGTGGCTGGGTT-CTGGGACGGGCCCGCCC 930

Query:   550 ACCAGTGATGCTCCTGGGACCCTACGCAATCTGCGCCTGCGTCTCATCAGTCGCCCCACA 609
             |  ||  | ||  | |   |||| | ||  | | ||| | | | | ||  | |  ||  |
Sbjct:   931 AGGAGC-AGGCATCAGCT-CCCT-C-CAGACAGTGCCCGTGACCC-TCC-TGGGGCCT-A 983

Query:   610 TGTAACTGTATCTACAACCA-GCTGCACCAGCGACACCTGTCCAACCCGGCCCGGCCTGG 668
```

```
                    |   ||| | |      || || || || | |     || | | |||    | ||| ||
Sbjct:    984 GGGC-CTGCAGCCGGCTGCATGCAGCTCCTGGGGTGATGGC-AGCCCTATTCTGCCGGG 1041

Query:    669 GATGCTATGTGGGGGCCCCCAGCCTGGG-GTGCAGGGCCCCTGTCAGGGAGATTCCGGGG 727
              |||| | |||    |  |   |  |||| | || |  |  |||| ||||    || ||||
Sbjct:   1042 GATGGTGTGTACCAGTGCT--G--TGGGTGAGCTGCCCAGCTGTGAGGGCCTGTCTGGGG 1097

Query:    728 G-CC-CTG-TGC-TGTGCCTCGAGCCTGACGGACACTGGGTTCAGGCTGG-CATCATCAG 782
                || ||| ||| || |    | | |    ||    |||   | |||||  || | || |
Sbjct:   1098 CACCACTGGTGCATGAGGTGAGGGGCACATGGTTCCTGGC--CGGGCTGCACAGCTTCGG 1155

Query:    783 CTTTGCAT-C-AAG-CTGTGCCCAGGAGGACG--CT-C-CTGTGCTGC-TGACCAACACA 834
                 |||   | ||| |    ||  ||  | ||   | | |   ||| | || | |  |
Sbjct:   1156 AGATGCTTGCCAAGGCCCCGCC-AGGCCGGCGGTCTTCACCGCGCTCCCTGCCTATGAGG 1214

Query:    835 GCTGC-TCA-CAGTTCCTGG-CTG-CAGGCTCGAGTTC 868
               |||  ||| |||||  ||| ||| |||| || | |||
Sbjct:   1215 ACTGGGTCAGCAGTT--TGGACTGGCAGG-TCTACTTC 1249
```

Figure 10. BLASTP identity search for the protein of the invention.

```
>patp:Y41704 Human PRO351 protein sequence - Homo sapiens, 571 aa. (SEQ ID
NO:65)

         Length = 571

  Plus Strand HSPs:

 Score = 2544 (895.5 bits), Expect = 1.1e-263, P = 1.1e-263
 Identities = 476/493 (96%), Positives = 479/493 (97%), Frame = +1

Query:     19 MLLSSLVSLAGSVYLAWILFFVLYDFCIVCITTYAINVSLMWLSFRKVQEPQGQPKPQEG 198
              |||||||||||||||||||||||||||||||||||||||||||||||||||||+ | + |
Sbjct:      1 MLLSSLVSLAGSVYLAWILFFVLYDFCIVCITTYAINVSLMWLSFRKVQEPQGKAK-RHG 59

Query:    199 NTVPGEWPWQASVRRQGAHICSGSLVADTWVLTAAHCFEKAAATELNS--CVRDS----- 357
              ||||||||||||||||||||||||||||||||||||||||||||||||   |  |
Sbjct:     60 NTVPGEWPWQASVRRQGAHICSGSLVADTWVLTAAHCFEKAAATELNSWSVVLGSLQREG 119

Query:    358 -APGAEEVGVAALQLPRAYNHYSQGSDLALLQLAHPTTHTPLCLPQPAHRFPFGASCWAT 534
              +|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    120 LSPGAEEVGVAALQLPRAYNHYSQGSDLALLQLAHPTTHTPLCLPQPAHRFPFGASCWAT 179

Query:    535 GWDQDTSDAPGTLRNLRLRLISRPTCNCIYNQLHQRHLSNPARPGMLCGGPQPGVQGPCQ 714
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    180 GWDQDTSDAPGTLRNLRLRLISRPTCNCIYNQLHQRHLSNPARPGMLCGGPQPGVQGPCQ 239

Query:    715 GDSGGPVLCLEPDGHWVQAGIISFASSCAQEDAPVLLTNTAAHSSWLQARVQGAAFLAQS 894
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    240 GDSGGPVLCLEPDGHWVQAGIISFASSCAQEDAPVLLTNTAAHSSWLQARVQGAAFLAQS 299

Query:    895 PETPEMSDEDSCVACGSLRTAGPQAGAPSPWPWEARLMHQGQLACGGALVSEEAVLTAAH 1074
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    300 PETPEMSDEDSCVACGSLRTAGPQAGAPSPWPWEARLMHQGQLACGGALVSEEAVLTAAH 359

Query:   1075 CFIGRQAPEEWSVGLGTRPEEWGLKQLILHGAYTHPEGGYDMALLLLAQPVTLGASLRPL 1254
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    360 CFIGRQAPEEWSVGLGTRPEEWGLKQLILHGAYTHPEGGYDMALLLLAQPVTLGASLRPL 419

Query:   1255 CLPYADHHLPDGERGWVLGRARPGAGISSLQTVPVTLLGPRACSRLHAAPGGDGSPILPG 1434
              |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    420 CLPYPDHHLPDGERGWVLGRARPGAGISSLQTVPVTLLGPRACSRLHAAPGGDGSPILPG 479

Query:   1435 MVCTSAVGELPSCE 1476
              ||||||||||||||
Sbjct:    480 MVCTSAVGELPSCE 493
```

```
 Score = 324 (114.1 bits), Expect = 7.0e-26, P = 7.0e-26 (SEQ ID NO:111)
 Identities = 91/250 (36%), Positives = 123/250 (49%), Frame = +1

Query:   187 PQEGNTVPGEWPWQASVRRQGAHICSGSLVADTWVLTAAHCFEKAAATELNSCVRDSAPG 366
             || |    |   |||+| +  ||    | |+||++ ||||||||    | |  |     + |
Sbjct:   322 PQAG--APSPWPWEARLMHQGQLACGGALVSEEAVLTAAHCFIGRQAPEEWSVGLGTRP- 378

Query:   367 AEEVGVAALQLPRAYNHYSQGSDLALLQLAHPTTH----TPLCLPQPAHRFPFGASCWAT 534
              || |+  | |  || |   | |+||| || | |       ||||| | |  | |   |
Sbjct:   379 -EEWGLKQLILHGAYTHPEGGYDMALLLLAQPVTLGASLRPLCLPYPDHHLPDGERGWVL 437

Query:   535 GWDQDTSDAPGTLRNLRLRLISRPTCNCIYNQLHQRHLSN--PARPGMLCGGPQPGVQGP 708
             |  +  +    +|+ + + |+    |+    +||     +  |  |||+|      |
Sbjct:   438 GRARPGAGI-SSLQTVPVTLLGPRACS----RLHAAPGGDGSPILPGMVCTSAV-GELPS 491

Query:   709 CQGDSGGPVLCLEPDGHWVQAGIISFASSCAQEDAPVLLTNTAAHSSWLQARVQGAAFLA 888
             |+| || | |   |  | |  ||+ ||  +|      | + |   |+  |+ + +     + |
Sbjct:   492 CEGLSGAP-LVHEVRGTWFLAGLHSFGDACQGPARPAVFTALPAYEDWVSS-LDWQVYFA 549

Query:   889 QSPETPEMSDEDSCVA 936
             + || || ++  ||+|
Sbjct:   550 EEPE-PE-AEPGSCLA 563
```

>patp:Y90291 Human peptidase, HPEP-8 protein sequence - Homo sapiens, 267 aa. (SEQ ID NO:66)

```
            Length = 267

  Plus Strand HSPs:

 Score = 1028 (361.9 bits), Expect = 5.0e-103, P = 5.0e-103
 Identities = 189/189 (100%), Positives = 189/189 (100%), Frame = +1

Query:   910 MSDEDSCVACGSLRTAGPQAGAPSPWPWEARLMHQGQLACGGALVSEEAVLTAAHCFIGR 1089
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:     1 MSDEDSCVACGSLRTAGPQAGAPSPWPWEARLMHQGQLACGGALVSEEAVLTAAHCFIGR 60

Query:  1090 QAPEEWSVGLGTRPEEWGLKQLILHGAYTHPEGGYDMALLLLAQPVTLGASLRPLCLPYA 1269
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    61 QAPEEWSVGLGTRPEEWGLKQLILHGAYTHPEGGYDMALLLLAQPVTLGASLRPLCLPYA 120

Query:  1270 DHHLPDGERGWVLGRARPGAGISSLQTVPVTLLGPRACSRLHAAPGGDGSPILPGMVCTS 1449
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   121 DHHLPDGERGWVLGRARPGAGISSLQTVPVTLLGPRACSRLHAAPGGDGSPILPGMVCTS 180

Query:  1450 AVGELPSCE 1476
             |||||||||
Sbjct:   181 AVGELPSCE 189

 Score = 316 (111.2 bits), Expect = 4.2e-27, P = 4.2e-27 (SEQ ID NO:112)
 Identities = 90/250 (36%), Positives = 122/250 (48%), Frame = +1

Query:   187 PQEGNTVPGEWPWQASVRRQGAHICSGSLVADTWVLTAAHCFEKAAATELNSCVRDSAPG 366
             || |    |   |||+| +  ||    | |+||++ ||||||||    | |  |     + |
Sbjct:    18 PQAG--APSPWPWEARLMHQGQLACGGALVSEEAVLTAAHCFIGRQAPEEWSVGLGTRP- 74

Query:   367 AEEVGVAALQLPRAYNHYSQGSDLALLQLAHPTTH----TPLCLPQPAHRFPFGASCWAT 534
              || |+  | |  || |   | |+||| || | |       |||||   | | |   |
Sbjct:    75 -EEWGLKQLILHGAYTHPEGGYDMALLLLAQPVTLGASLRPLCLPYADHHLPDGERGWVL 133

Query:   535 GWDQDTSDAPGTLRNLRLRLISRPTCNCIYNQLHQRHLSN--PARPGMLCGGPQPGVQGP 708
             |  +  +    +|+ + + |+    |+    +||     +  |  |||+|      |
Sbjct:   134 GRARPGAGI-SSLQTVPVTLLGPRACS----RLHAAPGGDGSPILPGMVCTSAV-GELPS 187

Query:   709 CQGDSGGPVLCLEPDGHWVQAGIISFASSCAQEDAPVLLTNTAAHSSWLQARVQGAAFLA 888
             |+| || | |   |  | |  ||+ ||  +|      | + |   |+  |+ + +     + |
```

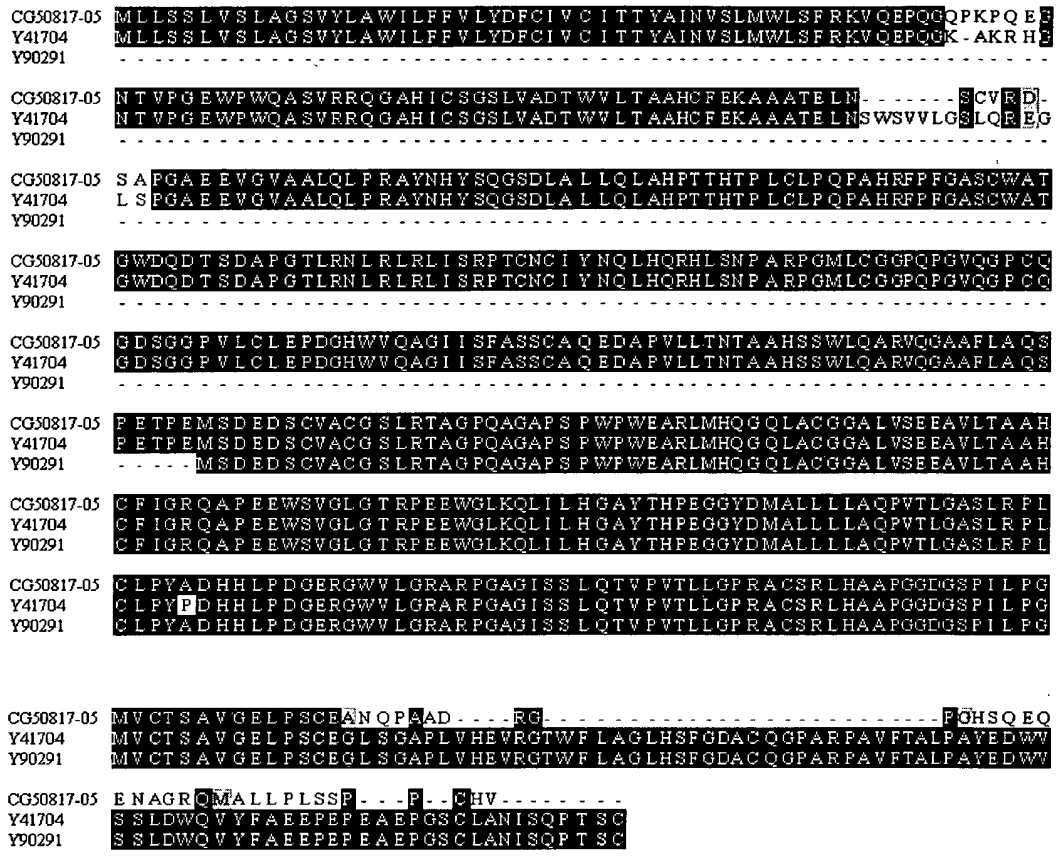
```
Sbjct:   188 CEGLSGAP-LVHEVRGTWFLAGLHSFGDACQGPARPAVFTALPAYEDWVSS-LDWQVYFA 245

Query:   889 QSPETPEMSDEDSCVA 936
             + || || ++  ||+|
Sbjct:   246 EEPE-PE-AEPGSCLA 259
```

Table 11. BLASTN identity search (versus the human SeqCalling database for the Peptidase-like protein of the invention.

```
>s3aq:153687026 Category D: 377 frag (6 5'sig-CG, 204 non-5'sig-CG, 167 non-CG (SEQ ID NO:67)
            EST), 1114 bp.
            Length = 1114

  Minus Strand HSPs:

 Score = 894 (134.1 bits), Expect = 3.1e-35, P = 3.1e-35
 Identities = 182/186 (97%), Positives = 182/186 (97%), Strand = Minus / Plus

Query:   186 CTTGGGTTGGCCCTGGGGTTCTTGGACCTTCCGGAAACTGAGCCACATCAGGCTCACGTT 127
             ||| |  | ||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   413 CTTAGCCTTGCCCTGGGGTTCTTGGACCTTCCGGAAACTGAGCCACATCAGGCTCACGTT 472

Query:   126 GATAGCATAGGTGGTGATACAAACAATGCAGAAATCATAGAGCACGAAGAACAGGATCCA 67
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   473 GATAGCATAGGTGGTGATACAAACAATGCAGAAATCATAGAGCACGAAGAACAGGATCCA 532

Query:    66 GGCCAGGTAGACAGAACCAGCGAGAGACACCAGGGAGCTCAGCAGCATCAGGACAGAGGC 7
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   533 GGCCAGGTAGACAGAACCAGCGAGAGACACCAGGGAGCTCAGCAGCATCAGGACAGAGGC 592

Query:     6 CCAGCG 1
             ||||||
Sbjct:   593 CCAGCG 598


>s3aq:152507187 17 frag (1 5'sig-CG, 7 non-5'sig-CG, 9 non-CG EST), 588 bp. (SEQ ID NO:68)
            Length = 588

  Plus Strand HSPs:

 Score = 882 (132.3 bits), Expect = 2.1e-34, P = 2.1e-34
 Identities = 178/180 (98%), Positives = 178/180 (98%), Strand = Plus / Plus

Query:     1 CGCTGGGCCTCTGTCCTGATGCTGCTGAGCTCCCTGGTGTCTCTCGCTGGTTCTGTCTAC 60
             ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Sbjct:   367 CGCTGGGCCTCTGTCCTGATGCTGCTGAGCTCCCTGGTGTCTCTCGCTGTTTCTGTCTAC 426

Query:    61 CTGGCCTGGATCCTGTTCTTCGTGCTCTATGATTTCTGCATTGTTTGTATCACCACCTAT 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   427 CTGGCCTGGATCCTGTTCTTCGTGCTCTATGATTTCTGCATTGTTTGTATCACCACCTAT 486

Query:   121 GCTATCAACGTGAGCCTGATGTGGCTCAGTTTCCGGAAGGTCCAAGAACCCCAGGGCCAA 180
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Sbjct:   487 GCTATCAACGTGAGCCTGATGTGGCTCAGTTTCCGGAAGGTCCAAGAACCCCAGGGGCAA 546


>s3aq:153485867 Category D: 3 frag (1 non-5'sig-CG, 2 non-CG EST), 612 bp. (SEQ ID NO:69)
            Length = 612

  Plus Strand HSPs:

 Score = 785 (117.8 bits), Expect = 1.7e-29, P = 1.7e-29
```

```
 Identities = 157/157 (100%), Positives = 157/157 (100%), Strand = Plus / Plus

Query:     1 CGCTGGGCCTCTGTCCTGATGCTGCTGAGCTCCCTGGTGTCTCTCGCTGGTTCTGTCTAC 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   456 CGCTGGGCCTCTGTCCTGATGCTGCTGAGCTCCCTGGTGTCTCTCGCTGGTTCTGTCTAC 515

Query:    61 CTGGCCTGGATCCTGTTCTTCGTGCTCTATGATTTCTGCATTGTTTGTATCACCACCTAT 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   516 CTGGCCTGGATCCTGTTCTTCGTGCTCTATGATTTCTGCATTGTTTGTATCACCACCTAT 575

Query:   121 GCTATCAACGTGAGCCTGATGTGGCTCAGTTTCCGGA 157
             |||||||||||||||||||||||||||||||||||||
Sbjct:   576 GCTATCAACGTGAGCCTGATGTGGCTCAGTTTCCGGA 612

>s3aq:153485864 Category D: 2 frag (2 non-5'sig-CG), 425 bp. (SEQ ID NO:70)
            Length = 425

  Plus Strand HSPs:

 Score = 785 (117.8 bits), Expect = 2.4e-29, P = 2.4e-29
 Identities = 157/157 (100%), Positives = 157/157 (100%), Strand = Plus / Plus

Query:     1 CGCTGGGCCTCTGTCCTGATGCTGCTGAGCTCCCTGGTGTCTCTCGCTGGTTCTGTCTAC 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   269 CGCTGGGCCTCTGTCCTGATGCTGCTGAGCTCCCTGGTGTCTCTCGCTGGTTCTGTCTAC 328

Query:    61 CTGGCCTGGATCCTGTTCTTCGTGCTCTATGATTTCTGCATTGTTTGTATCACCACCTAT 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   329 CTGGCCTGGATCCTGTTCTTCGTGCTCTATGATTTCTGCATTGTTTGTATCACCACCTAT 388

Query:   121 GCTATCAACGTGAGCCTGATGTGGCTCAGTTTCCGGA 157
             |||||||||||||||||||||||||||||||||||||
Sbjct:   389 GCTATCAACGTGAGCCTGATGTGGCTCAGTTTCCGGA 425
```

Table 12. ClustalW alignment of the protein of the invention.



Information for the ClustalW proteins:

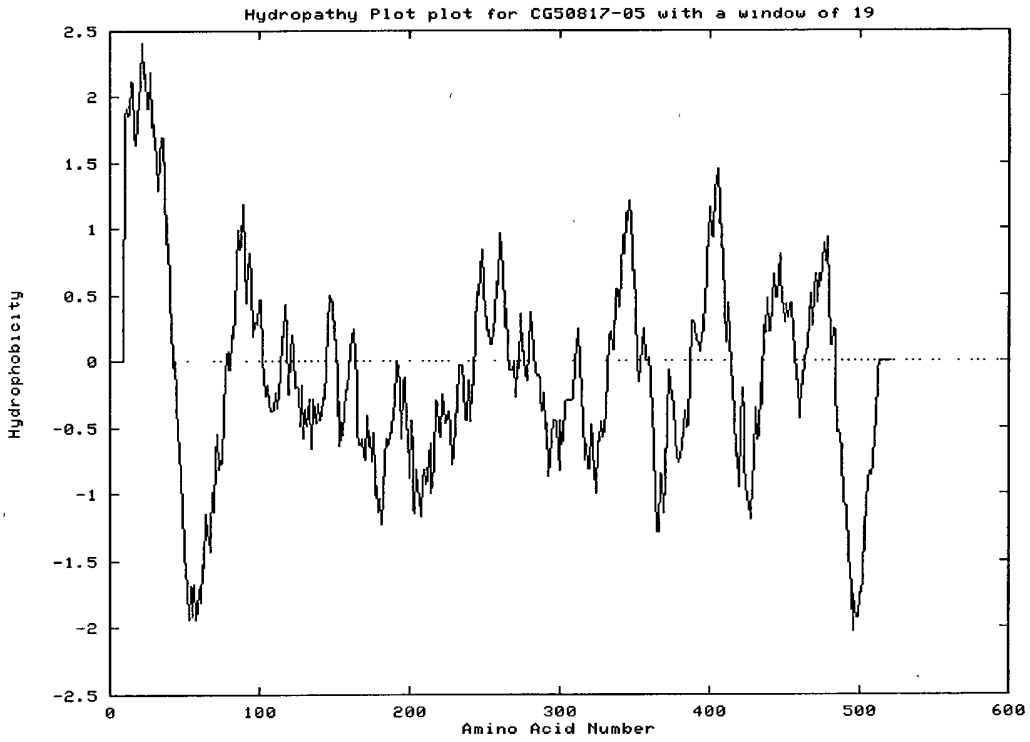
Accno	Common Name	Length
CG50817-05 **(SEQ ID NO:45)**	novel Peptidase-like protein	
Y41704 **(SEQ ID NO:122)**	Human PRO351 protein sequence.	571
Y90291**(SEQ ID NO:123)**	Human peptidase, HPEP-8 protein sequence.	267

In the alignment shown above, black outlined amino acid residues indicate regions of conserved sequence (i.e., regions that may be required to preserve structural or functional properties); greyed amino acid residues can be mutated to a residue with comparable steric and/or chemical properties without altering protein structure or function (e.g. L to V, I, or M);

non-highlighted amino acid residues can potentially be mutated to a much broader extent without altering structure or function.

Table 13. Psort, Signal P and hydropathy results for CG50817-05

```
                    plasma membrane --- Certainty=0.6850(Affirmative) < succ>
 endoplasmic reticulum (membrane) --- Certainty=0.6400(Affirmative) < succ>
                      Golgi body --- Certainty=0.3700(Affirmative) < succ>
           microbody (peroxisome) --- Certainty=0.1187(Affirmative) < succ>

 INTEGRAL Likelihood = -8.44 Transmembrane 15 - 31 (1 - 38)

 Seems to be a Type II (Ncyt Cexo) membrane protein
Is the sequence a signal peptide?
# Measure  Position  Value  Cutoff  Conclusion
  max. C    36       0.688   0.37   YES
  max. Y    36       0.555   0.34   YES
  max. S    10       0.991   0.88   YES
  mean S     1-35    0.875   0.48   YES
# Most likely cleavage site between pos. 35 and 36: TYA-IN
```



SECP 13

A SECP13 nucleic acid and polypeptide according to the invention includes the nucleic acid sequence (SEQ ID NO:46) and encoded polypeptide sequence (SEQ ID NO:47) of clone CG50817-06 directed toward novel peptidase (HPEP-8)-like proteins and nucleic acids encoding them. This is a related variant of SECP11 and SECP12, clones CG50817-04 and CG50817-05. Figure 18 illustrates the nucleic acid sequence and amino acid sequences respectively. This clone includes a nucleotide sequence (SEQ ID NO:46) of 1200 bp. The nucleotide sequence includes an open reading frame (ORF) beginning with an ATG initiation codon at nucleotides 33-35 and ending with a TGA codon at nucleotides 945-947. Putative untranslated regions, if any, are found upstream from the initiation codon and downstream from the termination codon. The encoded protein having 304 amino acid residues is presented using the one-letter code in Figure 18.

The protein encoded by clone CG50817-06 is predicted by the PSORT program to the cytoplasm with a certainty of 0.4500, and does not appear to be a signal protein (see Table 18 below).

The DNA sequence and protein sequence for a novel Peptidase-like gene or one of its splice forms thus derived is reported here as the invention CG50817-06. The Genomic clones having regions with 100% identity to the extended sequence thus obtained were identified by BLASTN searches with the extended sequence against human genomic databases. The genomic clone was selected for further analysis because this identity indicates that these clones contain the genomic locus for these SeqCalling assemblies.

The regions defined by all approaches were then manually integrated and manually corrected for apparent inconsistencies that may have arisen, for example, from miscalled bases in the original fragments used, or from discrepancies between predicted homolgy to a protein of similarity to derive the final sequence of the invention CG50817-06 reported here. When necessary, the process to identify and analyze SeqCalling assemblies, ESTs and genomic clones was reiterated to derive the full length sequence.

Similarities

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 840 of 842 bases (99%) identical to a gb:Z34002 Human PRO351 nucleotide sequence from Homo sapiens (Tables 14 and 16). The full amino acid sequence of

the protein of the invention was found to have 278 of 279 amino acid residues (99%) identical to, and 278 of 279 amino acid residues (99%) similar to, the 571 amino acid residue Y41704 Human PRO351 protein from Homo sapiens (Table 15).

A multiple sequence alignment is given in Table 17, with the protein of the invention being shown on the first line in a ClustalW analysis comparing the protein of the invention with related protein sequences.

The presence of identifiable domains in the protein disclosed herein was determined by searches using algorithms such as PROSITE, Blocks, Pfam, ProDomain, Prints and then determining the Interpro number by crossing the domain match (or numbers) using the Interpro website. The results indicate that this protein contains the following protein domains (as defined by Interpro) at the indicated positions: domain name trypsin at amino acid positions 1 to 62, domain name trypsin at amino acid positions 95 to 259. This indicates that the sequence of the invention has properties similar to those of other proteins known to contain this/these domain(s) and similar to the properties of these domains.

Chromosomal information:

The Peptidase disclosed in this invention maps to chromosome 16. This information was assigned using OMIM, the electronic northern bioinformatic tool implemented by CuraGen Corporation, public ESTs, public literature references and/or genomic clone homologies. This was executed to derive the chromosomal mapping of the SeqCalling assemblies, Genomic clones, literature references and/or EST sequences that were included in the invention.

Tissue expression

The Peptidase disclosed in this invention is expressed in at least the following tissues: Adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, and/or RACE sources.

Cellular Localization and Sorting

The SignalP, Psort and/or Hydropathy profile for the Peptidase-like protein are shown in Table 18. The results predict that this sequence has no signal peptide and is likely to be localized in the cytoplasm with a certainty of 0.4500 predicted by PSORT.

Functional Variants and Homologs

The novel nucleic acid of the invention encoding a Peptidase-like protein includes the nucleic acid whose sequence is provided in Figure 18, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Figure 18 while still encoding a protein that maintains its Peptidase-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 1% of the residues may be so changed.

The novel protein of the invention includes the Peptidase-like protein whose sequence is provided in Figure 18. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Figure 18 while still encoding a protein that maintains its Peptidase-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 1% of the bases may be so changed.

Antibodies

The invention further encompasses antibodies and antibody fragments, such as Fab, (Fab)2 or single chain FV constructs, that bind immunospecifically to any of the proteins of the

invention. Also encompassed within the invention are peptides and polypeptides comprising sequences having high binding affinity for any of the proteins of the invention, including such peptides and polypeptides that are fused to any carrier partcle (or biologically expressed on the surface of a carrier) such as a bacteriophage particle.

Uses of the Compositions of the Invention

The protein similarity information, expression pattern, and map location for the Peptidase-like protein and nucleic acid disclosed herein suggest that this Peptidase may have important structural and/or physiological functions characteristic of the Serine protease family. Therefore, the nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications and as a research tool. These include serving as a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed, as well as potential therapeutic applications such as the following: (i) a protein therapeutic, (ii) a small molecule drug target, (iii) an antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) a nucleic acid useful in gene therapy (gene delivery/gene ablation), and (v) a composition promoting tissue regeneration in vitro and in vivo (vi) biological defense weapon.

The nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications implicated in various diseases and disorders described below and/or other pathologies. For example, the compositions of the present invention will have efficacy for treatment of patients suffering from: cell proliferative disorder; arteriosclerosis; psoriasis; myelofibrosis; cancer; autoimmune disorder; Crohn's disease; inflammatory disorder; AIDS; anaemia; allergy; asthma; atherosclerosis; Grave's disease; multiple sclerosis; scleroderma; infection; diabetes; metabolic disorder; Addison's disease; cystic fibrosis; glycogen storage disease; obesity; nutritional edema, hypoproteinemia and other diseases, disorders and conditions of the like.

These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods.

Table 14. BLASTN identity search for the nucleic acid of the invention.

>patn:Z34002 Human PRO351 nucleotide sequence - Homo sapiens, 2365 bp. (SEQ ID NO:71)

Length = 2365

Plus Strand HSPs:

Score = 4192 (629.0 bits), Expect = 1.9e-184, P = 1.9e-184
Identities = 840/842 (99%), Positives = 840/842 (99%), Strand = Plus / Plus

```
Query:     1 AGCGACACCTGTCCAACCCGGCCCGGCCTGGGATGCTATGTGGGGGCCCCCAGCCTGGGG 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   936 AGCGACACCTGTCCAACCCGGCCCGGCCTGGGATGCTATGTGGGGGCCCCCAGCCTGGGG 995

Query:    61 TGCAGGGCCCCTGTCAGGGAGATTCCGGGGGCCCTGTGCTGTGCCTCGAGCCTGACGGAC 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   996 TGCAGGGCCCCTGTCAGGGAGATTCCGGGGGCCCTGTGCTGTGCCTCGAGCCTGACGGAC 1055

Query:   121 ACTGGGTTCAGGCTGGCATCATCAGCTTTGCATCAAGCTGTGCCCAGGAGGACGCTCCTG 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1056 ACTGGGTTCAGGCTGGCATCATCAGCTTTGCATCAAGCTGTGCCCAGGAGGACGCTCCTG 1115

Query:   181 TGCTGCTGACCAACACAGCTGCTCACAGTTCCTGGCTGCAGGCTCGAGTTCAGGGGGCAG 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1116 TGCTGCTGACCAACACAGCTGCTCACAGTTCCTGGCTGCAGGCTCGAGTTCAGGGGGCAG 1175

Query:   241 CTTTCCTGGCCCAGAGCCCAGAGACCCCGGAGATGAGTGATGAGGACAGCTGTGTAGCCT 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1176 CTTTCCTGGCCCAGAGCCCAGAGACCCCGGAGATGAGTGATGAGGACAGCTGTGTAGCCT 1235

Query:   301 GTGGATCCTTGAGGACAGCAGGTCCCCAGGCAGGAGCACCCTCCCCATGGCCCTGGGAGG 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1236 GTGGATCCTTGAGGACAGCAGGTCCCCAGGCAGGAGCACCCTCCCCATGGCCCTGGGAGG 1295

Query:   361 CCAGGCTGATGCACCAGGGACAGCTGGCCTGTGGCGGAGCCCTGGTGTCAGAGGAGGCGG 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1296 CCAGGCTGATGCACCAGGGACAGCTGGCCTGTGGCGGAGCCCTGGTGTCAGAGGAGGCGG 1355

Query:   421 TGCTAACTGCTGCCCACTGCTTCATTGGGCGCCAGGCCCCAGAGGAATGGAGCGTAGGGC 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1356 TGCTAACTGCTGCCCACTGCTTCATTGGGCGCCAGGCCCCAGAGGAATGGAGCGTAGGGC 1415

Query:   481 TGGGGACCAGACCGGAGGAGTGGGGCCTGAAGCAGCTCATCCTGCATGGAGCCTACACCC 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1416 TGGGGACCAGACCGGAGGAGTGGGGCCTGAAGCAGCTCATCCTGCATGGAGCCTACACCC 1475

Query:   541 ACCCTGAGGGGGGCTACGACATGGCCCTCCTGCTGCTGGCCCAGCCTGTGACACTGGGAG 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1476 ACCCTGAGGGGGGCTACGACATGGCCCTCCTGCTGCTGGCCCAGCCTGTGACACTGGGAG 1535

Query:   601 CCAGCCTGCGGCCCCTCTGCCTGCCCTATGCTGACCACCACCTGCCTGATGGGGAGCGTG 660
             |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Sbjct:  1536 CCAGCCTGCGGCCCCTCTGCCTGCCCTATCCTGACCACCACCTGCCTGATGGGGAGCGTG 1595

Query:   661 GCTGGGTTCTGGGACGGGCCCGCCCAGGAGCAGGCATCAGCTCCCTCCAGACAGTGCCCG 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1596 GCTGGGTTCTGGGACGGGCCCGCCCAGGAGCAGGCATCAGCTCCCTCCAGACAGTGCCCG 1655

Query:   721 TGACCCTCCTGGGGCCTAGGGCCTGCAGCCGGCTGCATGCAGCTCCTGGGGGTGATGGCA 780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1656 TGACCCTCCTGGGGCCTAGGGCCTGCAGCCGGCTGCATGCAGCTCCTGGGGGTGATGGCA 1715

Query:   781 GCCCTATTCTGCCGGGGATGGTGTGTACCAGTGCTGTGGGTGAGCTGCCCAGCTGTGAGG 840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1716 GCCCTATTCTGCCGGGGATGGTGTGTACCAGTGCTGTGGGTGAGCTGCCCAGCTGTGAGG 1775
```

```
Query:    841 CC 842
               |
Sbjct:   1776 GC 1777
```

Score = 1915 (287.3 bits), Expect = 1.4e-81, P = 1.4e-81 **(SEQ ID NO:114)**
Identities = 635/848 (74%), Positives = 635/848 (74%), Strand = Plus / Plus

```
Query:    353 CTGGGAGGCCAGGCTGATGCAC-CAGGGACAGCTGGCCTGTGGCGGAGC--CCTGG--TG 407
              |||   | |||| ||| || || | ||||  ||  ||||| |||    |  ||||   | 
Sbjct:   1508 CTGCTGGCCCAGCCTG-TG-ACACTGGGA--GCCAGCCTGCGGCCCCTCTGCCTGCCCTA 1563

Query:    408 TCA-GAGGAGGCGGTGC-TAACTGCTGCCCACTGCTTCATTGGGCGCCAGGCCC-CAGAG 464
              ||  ||  |     ||| | | ||  |  |   |||   ||  | | | |||||| |  ||
Sbjct:   1564 TCCTGACCACCACCTGCCTGA-TGGGGAGCGTGGCTGGGTTCTGGGACGGGCCCGCCCAG 1622

Query:    465 GAATGGAGCGTAGGGCTGGGACCAGACCGGAGGAGTGGGCCTGAAGCAGCTCAT--CC 522
              ||   | || |  | ||     |||||| |     |||   |||   |  ||  |   ||
Sbjct:   1623 GAGCAG-GCATCAG-CTCCCT-CCAGACAGTGCCCGTGAC-CCTCCTGGGGCCTAGGGCC 1678

Query:    523 TGCATGGAGCCTACACCCACCCTGAGGGGGCTACGACATGGCCCTCCTGCTGCTGGCCC 582
              |||| |  | || ||  || | |   |||||  | | ||  ||||| |  | |||| ||   
Sbjct:   1679 TGCA-GCCGGCTGCATGCAGC-TCCTGGGGGTGATGGCA--GCCCTATT-CTGCCGGGGA 1733

Query:    583 AGCCTGTG-ACACTGGGA-GCCAGCCTGCGGCCCCTCTGCCTGC-CCTATGCTGAC-CAC 638
               |  |||| || | | |  |   |   ||  ||||  |||   |  ||| | |||   |||
Sbjct:   1734 TGG-TGTGTAC-CAGTGCTGTGGGTGAGCTGCCCAGCTGTGAGGGCCTGT-CTGGGGCAC 1790

Query:    639 CACC--TGCCTGATGGGGAGCGTGGCTGGTTCTGGGACGGGCCCGCCCAGGAGCAGGCA 696
              |||   ||| ||| || |||  |   |  |||||  || |||||  || |||   | |  |
Sbjct:   1791 CACTGGTGCATGA-GGTGAGGGGCACATGGTTCCTGGCCGGGCT-GCACAGCTTCGGAGA 1848

Query:    697 T-CA-GCTCCCTCCA-GACAGTGCCCGTGACCCTCCTGGGGCCTAGGGCCTGCAGCCGGC 753
              | |  ||     ||  | ||| ||| | |  | ||   | ||     ||||   |  | |
Sbjct:   1849 TGCTTGCCAAGGCCCCGCCAG-GCCGGCGGTCTTCACCGCGCTCCCTGCCTAT-GAGGAC 1906

Query:    754 TGCATGCAGCTCCTGGGGGTGATGGCAGCCCTA-TTCTGCCGGGGATGGTGTGTACCAGT 812
              ||  | ||||   | ||  ||  | |||  ||| ||| |||||| |||    | |  | || 
Sbjct:   1907 TGGGT-CAGCAGTTTGGACTG--G-CAGGTCTACTTC-GCCGAGGAACCAGAGCCCGAG- 1960

Query:    813 GCTGTGGGTG-A-GCTGCCCAGCTGTGAG--GCCAACCAACCAGCTGCTGACAGGGGACC 868
              |||| |  || | ||||||  ||    |   |||||||||||||||||||||||||||||
Sbjct:   1961 GCTGAGCCTGGAAGCTGCCTGGCCAACATAAGCCAACCAACCAGCTGCTGACAGGGGACC 2020

Query:    869 TGGCCATTCTCAGGAACAAGAGAATGCAGGCAGGCAAATGGCATTACTGCCCCTGTCCTC 928
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   2021 TGGCCATTCTCAGGA-CAAGAGAATGCAGGCAGGCAAATGGCATTACTGCCCCTGTCCTC 2079

Query:    929 CCCACCCTGTCATGTGTGATTCCAGGCACCAGGGCAGGCCCAGAAGCCCAGCAGCTGTGG 988
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   2080 CCCACCCTGTCATGTGTGATTCCAGGCACCAGGGCAGGCCCAGAAGCCCAGCAGCTGTGG 2139

Query:    989 GAAGGAACCTGCCTGGGGCCACAGGTGCCCACTCCCCACCCTGCAGGACAGGGGTGTCTG 1048
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   2140 GAAGGAACCTGCCTGGGGCCACAGGTGCCCACTCCCCACCCTGCAGGACAGGGGTGTCTG 2199

Query:   1049 TGGACACTCCCACACCCAACTCTGCTACCAAGCAGGCGTCTCAGCTTTCCTCCTCCTTTA 1108
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   2200 TGGACACTCCCACACCCAACTCTGCTACCAAGCAGGCGTCTCAGCTTTCCTCCTCCTTTA 2259

Query:   1109 CCCTTTCAGATACAATCACGCCAGCCACGTTGTTTGAAAATTTCTTTTTTTGGGGGGCA 1168
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   2260 CTCTTTCAGATACAATCACGCCAGCCACGTTGTTTGAAAATTTCTTTTTTTGGGGGGCA 2319

Query:   1169 GCAGTTTTCCTTTTTTTAAACTTAAATAAATT 1200
              ||||||||||||||||||||||||||||||||
Sbjct:   2320 GCAGTTTTCCTTTTTTTAAACTTAAATAAATT 2351
```

Score = 267 (40.1 bits), Expect = 0.0078, P = 0.0078 **(SEQ ID NO:115)**

```
 Identities = 349/598 (58%), Positives = 349/598 (58%), Strand = Plus / Plus

Query:   275 GAGTGA-TGAGGACAGCTGTGTAGCCTGTGGATCCTTGAGGACAGCAGGTCCCCAGGCAG 333
             | |||  || ||||||| |||  |||   ||  ||   | | |  ||||    ||  ||
Sbjct:   424 GCGTGCCTGTGGACAGC-GTG--GCCCC-GGCCCCCCCAAGCCT-CAGGAGGGCAA-CAC 477

Query:   334 GAGCACCCTCCCCA-TGGCCCTGGGAGGCCAGGCTGATGCACCAGGGACAGCTGGCCTGT 392
              ||  ||||  | | |||||||||| |||||||  |||  |   || |||   |    |||
Sbjct:   478 -AGT-CCCTGGCGAGTGGCCCTGGCAGGCCAGTGTGAGGAGGCAAGGAGCCCACATCTGC 535

Query:   393 GGCGGAGCCCTGGTGTCAGAGGAGGCGGTGCTAACTGCTGCCCACTGCTTC-ATTGGGCG 451
              ||||  |||||||| ||||      ||| || ||||||||||||||||  |   |||
Sbjct:   536 AGCGGCTCCCTGGTGGCAGACACCTGGGTCCTCACTGCTGCCCACTGCTTTGAAAAGGCA 595

Query:   452 CCAGGCCCCAGAG--GAATGGAGCGT-AG-GG-CTGGGGACCAGACCGGAGGAGTG-GGG 505
              ||| |  ||||   ||||      || || || |  |||  |   | | ||  ||| |||
Sbjct:   596 GCAG-CAACAGAACTGAATTCCTGGTCAGTGGTCCTGGGTTCT--CTGCAGC-GTGAGGG 651

Query:   506 CCTGAAGCAGCTCATCCTGCAT-GGAGCCTACACCCACCCTGAGGGGGGCTACGAC--AT 562
             || | ||  ||    | | |  || |       |  |||||  |  | | | | | |  ||
Sbjct:   652 ACTCA-GCC-CTGGGGCCGAAGAGGTGGGGGTGGCTGCCCTGCAGTTGCCCAGGGCCTAT 709

Query:   563 GGCC-CTCCTGCTGCTGGCCCAG-CCTGTGACACTGGGAGCCAGCCTGCGGCCCCTCTGC 620
               || || | ||    ||| ||| |||| | | |||   || ||| | || || | |  |
Sbjct:   710 AACCACTACAGCCAG-GGCTCAGACCTG-GCC-CTGCT-GC-AGC-T-CGCCCACCCCAC 762

Query:   621 CTGCCCTATGCTGACCACCACCTGCCTGATGGGGAGCGTGGCTGGGT-TCTGG-GACGG- 677
               ||| |  |   || |  ||||||  |    |  | | |||     | ||| | |
Sbjct:   763 GA-CCC-ACACACCCCTCTGCCTGCCCCAGCCCGCCCATCGCTTCCCCTTTGGAGCCTCC 820

Query:   678 -GCCCGCCCAGGAGCAGGCATCAGCTCCCTCCAGACAGTGC-CC-GTGACCCTCCTGGGG 734
              ||  | |||   || || |||||  | |  |||  | ||| || | |||||| | |
Sbjct:   821 TGCTGGGCCACTGGCTGGGATCAGGACAC--CAGTGA-TGCTCCTGGGACCCTAC-GCAA 876

Query:   735 CCTAGGGCCTGCAGCCGGCTGCA-T-GCAGCTCCTGGGGGTG-ATGG-CAGCCCTATTCT 790
              || | |||||| | |   | || | ||  | | ||    || ||   || ||   | |
Sbjct:   877 TCT-GCGCCTGC-GTCTCAT-CAGTCGCCCCACATGTAACTGTATCTACAACCAGCTGCA 933

Query:   791 GCCGGGGATGG-TGTGTA-CCAGTGCTGTGGGTGAGCTGCCCAGCTGTGAGGCCAACCAA 848
              || | ||    |||  | || |  | | |  || | |||   | || | ||||  |||
Sbjct:   934 -CCAGCGACACCTGTCCAACCCGGCCCG-GCCTGGGATGCTATG-TGGG-GGCCC-CCAG 988

Query:   849 CCAGCTGCTGACAGGGGACCTGGC 872
              || |  | || |||||  |||| |
Sbjct:   989 CCTGGGG-TG-CAGGGCCCCTGTC 1010

>patn:A37664 Human peptidase, HPEP-8 coding sequence - Homo sapiens, 1661 bp (SEQ ID
NO:72)

           Length = 1661

  Plus Strand HSPs:

 Score = 3831 (574.8 bits), Expect = 5.6e-168, P = 5.6e-168
 Identities = 767/768 (99%), Positives = 767/768 (99%), Strand = Plus / Plus

Query:    75 CAGGGAGATTCCGGGGGCCCTGTGCTGTGCCTCGAGCCTGACGGACACTGGGTTCAGGCT 134
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   320 CAGGGAGATTCCGGGGGCCCTGTGCTGTGCCTCGAGCCTGACGGACACTGGGTTCAGGCT 379

Query:   135 GGCATCATCAGCTTTGCATCAAGCTGTGCCCAGGAGGACGCTCCTGTGCTGCTGACCAAC 194
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   380 GGCATCATCAGCTTTGCATCAAGCTGTGCCCAGGAGGACGCTCCTGTGCTGCTGACCAAC 439

Query:   195 ACAGCTGCTCACAGTTCCTGGCTGCAGGCTCGAGTTCAGGGGGCAGCTTTCCTGGCCCAG 254
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   440 ACAGCTGCTCACAGTTCCTGGCTGCAGGCTCGAGTTCAGGGGGCAGCTTTCCTGGCCCAG 499
```

```
Query:   255 AGCCCAGAGACCCCGGAGATGAGTGATGAGGACAGCTGTGTAGCCTGTGGATCCTTGAGG 314
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   500 AGCCCAGAGACCCCGGAGATGAGTGATGAGGACAGCTGTGTAGCCTGTGGATCCTTGAGG 559

Query:   315 ACAGCAGGTCCCCAGGCAGGAGCACCCTCCCCATGGCCCTGGGAGGCCAGGCTGATGCAC 374
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   560 ACAGCAGGTCCCCAGGCAGGAGCACCCTCCCCATGGCCCTGGGAGGCCAGGCTGATGCAC 619

Query:   375 CAGGGACAGCTGGCCTGTGGCGGAGCCCTGGTGTCAGAGGAGGCGGTGCTAACTGCTGCC 434
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   620 CAGGGACAGCTGGCCTGTGGCGGAGCCCTGGTGTCAGAGGAGGCGGTGCTAACTGCTGCC 679

Query:   435 CACTGCTTCATTGGGCGCCAGGCCCCAGAGGAATGGAGCGTAGGGCTGGGGACCAGACCG 494
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   680 CACTGCTTCATTGGGCGCCAGGCCCCAGAGGAATGGAGCGTAGGGCTGGGGACCAGACCG 739

Query:   495 GAGGAGTGGGGCCTGAAGCAGCTCATCCTGCATGGAGCCTACACCCACCCTGAGGGGGGC 554
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   740 GAGGAGTGGGGCCTGAAGCAGCTCATCCTGCATGGAGCCTACACCCACCCTGAGGGGGGC 799

Query:   555 TACGACATGGCCCTCCTGCTGCTGGCCCAGCCTGTGACACTGGGAGCCAGCCTGCGGCCC 614
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   800 TACGACATGGCCCTCCTGCTGCTGGCCCAGCCTGTGACACTGGGAGCCAGCCTGCGGCCC 859

Query:   615 CTCTGCCTGCCCTATGCTGACCACCACCTGCCTGATGGGGAGCGTGGCTGGGTTCTGGGA 674
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   860 CTCTGCCTGCCCTATGCTGACCACCACCTGCCTGATGGGGAGCGTGGCTGGGTTCTGGGA 919

Query:   675 CGGGCCCGCCCAGGAGCAGGCATCAGCTCCCTCCAGACAGTGCCCGTGACCCTCCTGGGG 734
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   920 CGGGCCCGCCCAGGAGCAGGCATCAGCTCCCTCCAGACAGTGCCCGTGACCCTCCTGGGG 979

Query:   735 CCTAGGGCCTGCAGCCGGCTGCATGCAGCTCCTGGGGGTGATGGCAGCCCTATTCTGCCG 794
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   980 CCTAGGGCCTGCAGCCGGCTGCATGCAGCTCCTGGGGGTGATGGCAGCCCTATTCTGCCG 1039

Query:   795 GGATGGTGTGTACCAGTGCTGTGGGTGAGCTGCCCAGCTGTGAGGCC 842
             |||||||||||||||||||||||||||||||||||||||||||||| |
Sbjct:  1040 GGATGGTGTGTACCAGTGCTGTGGGTGAGCTGCCCAGCTGTGAGGGC 1087

 Score = 1931 (289.7 bits), Expect = 3.7e-82, P = 3.7e-82 (SEQ ID NO:116)
 Identities = 635/848 (74%), Positives = 635/848 (74%), Strand = Plus / Plus

Query:   353 CTGGGAGGCCAGGCTGATGCAC-CAGGGACAGCTGGCCTGTGGCGGAGC--CCTGGTGTC 409
             |||   | |||| ||| || || | ||||  ||  ||||| |||    |  ||||   | 
Sbjct:   818 CTGCTGGCCCAGCCTG-TG-ACACTGGGA--GCCAGCCTGCGGCCCCTCTGCCTGCCCTA 873

Query:   410 AGAGGAGGCGGTGCTAACTGCTGCCCA-C-TG-CTTCATTGGGCGCCAGGCCC-CAGAGG 465
              |  ||       ||   || ||   | | || ||   ||  |   | ||||| |  |||
Sbjct:   874 TGCTGACCACCACCTGCCTGATGGGGAGCGTGGCTGGGTTCTGGGACGGGCCCGCCCAGG 933

Query:   466 AATGGAGCGTAGGGCTGGGGACCAGACCGGAGGAGTGGGGCCTGAAGCAGCTCAT--CCT 523
             |   | || |  | ||     |||||| |     |||   |||   |  ||  |   |||
Sbjct:   934 AGCAG-GCATCAG-CTCCCT-CCAGACAGTGCCCGTGAC-CCTCCTGGGGCCTAGGGCCT 989

Query:   524 GCATGGAGCCTACACCCACCCTGAGGGGGGCTACGACATGGCCCTCCTGCTGCTGGCCCA 583
             ||| |  | || ||  || | |   |||||  | | ||  |||||  | |||| ||    
Sbjct:   990 GCA-GCCGGCTGCATGCAGC-TCCTGGGGGTGATGGCA--GCCCTATT-CTGCCGGGGAT 1044

Query:   584 GCCTGTG-ACACTGGGA-GCCAGCCTGCGGCCCCTCTGCCTGC-CCTATGCTGAC-CACC 639
             |  |||| || | | |  |   |   || ||||  |||   |  ||| | |||   ||||
Sbjct:  1045 GG-TGTGTAC-CAGTGCTGTGGGTGAGCTGCCCAGCTGTGAGGGCCTGT-CTGGGGCACC 1101

Query:   640 ACC--TGCCTGATGGGGAGCGTGGCTGGGTTCTGGGACGGGCCCGCCCAGGAGCAGGCAT 697
             ||   ||| ||| || |||  |     |||||  || |||||  || |||   |    ||
Sbjct:  1102 ACTGGTGCATGA-GGTGAGGGGCACATGGTTCCTGGCCGGGCT-GCACAGCTTCGGAGAT 1159
```

```
Query:   698 -CA-GCTCCCTCCA-GACAGTGCCCGTGACCCTCCTGGGGCCTAGGGCCTGCAGCCGGCT 754
              |  ||     ||  | ||| ||| | |  | ||   | ||     ||||   |  | ||
Sbjct:  1160 GCTTGCCAAGGCCCCGCCAG-GCCGGCGGTCTTCACCGCGCTCCCTGCCTAT-GAGGACT 1217

Query:   755 GCATGCAGCTCCTGGGGGTGATGGCAGCCCTA-TTCTGCCGGGGATGGTGTGTACCAGTG 813
             |  | ||||   | ||  ||  | |||  ||| ||| |||| |||    | |  | || |
Sbjct:  1218 GGGT-CAGCAGTTTGGACTG--G-CAGGTCTACTTC-GCCGAGGAACCAGAGCCCGAG-G 1271

Query:   814 CTGTGGGTG-A-GCTGCCCAGCTGTGAG--GCCAACCAACCAGCTGCTGACAGGGGACCT 869
             ||| |  || | ||||||  ||    |   ||||||||||||||||||||||||||||||
Sbjct:  1272 CTGAGCCTGGAAGCTGCCTGGCCAACATAAGCCAACCAACCAGCTGCTGACAGGGGACCT 1331

Query:   870 GGCCATTCTCAGGAACAAGAGAATGCAGGCAGGCAAATGGCATTACTGCCCCTGTCCTCC 929
             |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1332 GGCCATTCTCAGGA-CAAGAGAATGCAGGCAGGCAAATGGCATTACTGCCCCTGTCCTCC 1390

Query:   930 CCACCCTGTCATGTGTGATTCCAGGCACCAGGGCAGGCCCAGAAGCCCAGCAGCTGTGGG 989
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1391 CCACCCTGTCATGTGTGATTCCAGGCACCAGGGCAGGCCCAGAAGCCCAGCAGCTGTGGG 1450

Query:   990 AAGGAACCTGCCTGGGGCCACAGGTGCCCACTCCCCACCCTGCAGGACAGGGGTGTCTGT 1049
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1451 AAGGAACCTGCCTGGGGCCACAGGTGCCCACTCCCCACCCTGCAGGACAGGGGTGTCTGT 1510

Query:  1050 GGACACTCCCACACCCAACTCTGCTACCAAGCAGGCGTCTCAGCTTTCCTCCTCCTTTAC 1109
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1511 GGACACTCCCACACCCAACTCTGCTACCAAGCAGGCGTCTCAGCTTTCCTCCTCCTTTAC 1570

Query:  1110 CCTTTCAGATACAATCACGCCAGCCACGTTGTTTTGAAAATTTCTTTTTTTGGGGGGCAG 1169
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1571 CCTTTCAGATACAATCACGCCAGCCACGTTGTTTTGAAAATTTCTTTTTTTGGGGGGCAG 1630

Query:  1170 CAGTTTTCCTTTTTTTAAACTTAAATAAATT 1200
             |||||||||||||||||||||||||||||||
Sbjct:  1631 CAGTTTTCCTTTTTTTAAACTTAAATAAATT 1661
```

Score = 559 (83.9 bits), Expect = 8.2e-17, P = 8.2e-17 **(SEQ ID NO:117)**
Identities = 609/1017 (59%), Positives = 609/1017 (59%), Strand = Plus / Plus

```
Query:     1 AGCGACACCTGTCCAACCCGGCCCGGCCTGGGATGCTATGTGGGGGCCCCCAGCCTGGGG 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    93 AGCGACACCTGTCCAACCCGGCCCGGCCTGGGATGCTATGTGGGGGCCCCCAGCCTGGGG 152

Query:    61 TGCAGGGCCCCTGTCAGGGA-GATTCCGGGG-GCCCTGT-GCTGTGCCTCGAGCCTGACG 117
             ||||||||||||||||||   |||   | |  |    |  || |      |||   | |
Sbjct:   153 TGCAGGGCCCCTGTCAGGTCTGATAGGGAGAAGAGAAGGAGCAGAAGGG-GAGGG-GCCT 210

Query:   118 GACACTGGGTTCAGGCTGGCA-TCATCAG--CTTTGCATCA-AGCTGTGCCCAGGAGGAC 173
              || ||||| |  || | | | ||| |||  ||  |    | |||||    ||| |||
Sbjct:   211 AACCCTGGGCTGGGGGTTGGACTCA-CAGGACTGGGGAAAGAGCTGCAATCAG-AGGGT 268

Query:   174 GCTCCTGTGCT-GCTGACCA-ACACAGCTGCTCACAGTTCCTGGCTGCA-GGCTC---G- 226
             | || ||   | ||||  |  |  || ||| ||   |  | | | |||  |||||   |
Sbjct:   269 G-TC-TGCCATAGCTGGGCTCAGGCATCTG-TCCTTGG-CTTTGTTGCCTGGCTCCAGGG 324

Query:   227 AG-TTCAGGGGGCAGCTTTCCTG-GCCCAGAGCCC-AGAGACCCCGGAGATGAGTGATGA 283
             || ||| ||||||  || | ||| |||  |||||  |  ||| | || | | || | ||
Sbjct:   325 AGATTCCGGGGGCC-CTGTGCTGTGCCTCGAGCCTGACGGACACTGG-GTTCAG-GCTG- 380

Query:   284 GGACAGCTGTGTAGCCTGTGGATCCT--TGAGGACAGCAGGTC-C-CCAG-GCAGGAGCA 338
              | || |  |  ||| | || |||    || |  ||| ||| | | || | || |  | |
Sbjct:   381 -G-CATCA-TC-AGCTT-TGCATCAAGCTGTGCCCAGGAGGACGCTCCTGTGCTGCTG-A 434

Query:   339 CCCTCCCCATGGCCCTGGGAGG-CCAGGCTG-ATGCACCAGGGACAGCTGGCCTGTGGCG 396
             ||  | | | |   ||   ||  || ||||| | || | ||   |||  |||   |  |
Sbjct:   435 CCAACAC-A-GCTGCTCACAGTTCCTGGCTGCAGGCTCGAGTT-CAGGGGGCAGCTTTCC 491

Query:   397 GAGCCCTGGTGTCAGAGGAGGCGGTGCTAACTGCTGCCCACTGCTTCATTGGGCGCCAGG 456
```

```
Sbjct:   492 TGGCCCAGAGCCCAGAGACCCCGGAGATGAGTGATGAGGACAGCTGTGTAGCCTGTG-GA 550

Query:   457 CCCCAGAGGAATGGAG--CGTAGGGCTGGGG-ACCAGACCGGAGGAGTGGGGCCTGAAGC 513
Sbjct:   551 TCCTTGAGGACAGCAGGTCCCCAGGCAGGAGCACCCTCCCCATGGCCCTGGGAGGCCAG- 609

Query:   514 AGCTCATCCTGCATGGAGC-CTACACCCACCCTGAGGGGGGCTA-C-GACATGGCCCTCC 570
Sbjct:   610 -GCTGATGCACCAGGGACAGCTGGCCTGTGGCGGAGCCCTGGTGTCAGAGGAGGCGGTGC 668

Query:   571 TG-CTGCTGGCCCAGCCTGTGACACTGGGAGCCAGCCTGCGGCCCCTCTGCCTGCCCTAT 629
Sbjct:   669 TAACTGCTG-CCCA--CTGCTTCATTGGGCGCCAGGCCCCAGAGGAA-TGGA-GCG-TAG 722

Query:   630 G-CTGACCACCAC-CTGCCTGA-TGGGGAGCGTGGCTGGGT-TCTGGGACGGGCCCGCCC 685
Sbjct:   723 GGCTGGGGACCAGACCGGAGGAGTGGGGCCTGAAGCAGCTCATCCTGCATGGAGCCTAC- 781

Query:   686 AGGAGCAGGCATCAGCTCC-CTCCAGACAGTGCCCGTGACCCTCCTGGG---GCCTAGGG 741
Sbjct:   782 ACC--CACCC-TGAGGGGGGCTAC-GACATGGCCC-TCCTGCTGCTGGCCCAGCCTGTGA 836

Query:   742 C-CTGC-AGCCGGC-TGCATGCAGCTCCTGGGGGTGATG-GCAG-CC-CTATTCTGCCGG 795
Sbjct:   837 CACTGGGAGCCAGCCTGCG-GCCCCTC-TGCCTGCCCTATGCTGACCACCAC-CTGCCTG 893

Query:   796 GGATGGTGTGTACCAGTGCTGTGGGT-GAGCT-GCCCAGCTGTGAGGCCAACCAACCAGC 853
Sbjct:   894 ATGGGGAGCGTGGCTGGGTTCTGGGACGGGCCCGCCCAGGAGC-AGGC--ATCAGCTCCC 950

Query:   854 TGCTGACAGGGGACCTGGCCATTCTCAGGAACAAGAGAATGCAGGCAGGCAA-ATGGCAT 912
Sbjct:   951 TCCAGACAGTGCCCGTGACCCTCCTGGGGC-CTAGGGCCTGCAGCC-GGCTGCATG-CAG 1007

Query:   913 -TACTGCCCCTG-TC-CTCCCC-ACCCTGTCATGTGTGATTCCAGGCACCAGGGCAGGCC 968
Sbjct:  1008 CTCCTGGGGGTGATGGCAGCCCTATTCTGCCG-G-G-GATGGTGTGTACCAGTGCTGTGG 1064

Query:   969 CAGAAGCCCAGCAGCTGTGGGAAGGAACCTGCCTGGGGC--CACAGGTGC 1016
Sbjct:  1065 GTGA-GCTGCCCAGCTGTGAG--GG--CCTGTCTGGGGCACCACTGGTGC 1109
```

Table 15. BLASTP identity search for the protein of the invention.

>patp:Y41704 Human PRO351 protein sequence - Homo sapiens, 571 aa. (SEQ ID NO:73)

Length = 571

Plus Strand HSPs:

Score = 1514 (533.0 bits), Expect = 1.6e-154, P = 1.6e-154
Identities = 278/279 (99%), Positives = 278/279 (99%), Frame = +3

```
Query:     3 RHLSNPARPGMLCGGPQPGVQGPCQGDSGGPVLCLEPDGHWVQAGIISFASSCAQEDAPV 182
Sbjct:   215 RHLSNPARPGMLCGGPQPGVQGPCQGDSGGPVLCLEPDGHWVQAGIISFASSCAQEDAPV 274

Query:   183 LLTNTAAHSSWLQARVQGAAFLAQSPETPEMSDEDSCVACGSLRTAGPQAGAPSPWPWEA 362
Sbjct:   275 LLTNTAAHSSWLQARVQGAAFLAQSPETPEMSDEDSCVACGSLRTAGPQAGAPSPWPWEA 334

Query:   363 RLMHQGQLACGGALVSEEAVLTAAHCFIGRQAPEEWSVGLGTRPEEWGLKQLILHGAYTH 542
```

```
Sbjct:   335 RLMHQGQLACGGALVSEEAVLTAAHCFIGRQAPEEWSVGLGTRPEEWGLKQLILHGAYTH 394

Query:   543 PEGGYDMALLLLAQPVTLGASLRPLCLPYADHHLPDGERGWVLGRARPGAGISSLQTVPV 722
             |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Sbjct:   395 PEGGYDMALLLLAQPVTLGASLRPLCLPYPDHHLPDGERGWVLGRARPGAGISSLQTVPV 454

Query:   723 TLLGPRACSRLHAAPGGDGSPILPGMVCTSAVGELPSCE 839
             |||||||||||||||||||||||||||||||||||||||
Sbjct:   455 TLLGPRACSRLHAAPGGDGSPILPGMVCTSAVGELPSCE 493

 Score = 225 (79.2 bits), Expect = 4.6e-15, P = 4.6e-15 (SEQ ID NO:118)
 Identities = 71/203 (34%), Positives = 95/203 (46%), Frame = +3

Query:   339 PSPWPWEARLMHQGQLACGGALVSEEAVLTAAHCFIGRQAPE--EWSVGLGT------RP 494
             |  |||+|  +   ||     |  |+||++  |||||||      | |    ||| ||+    |
Sbjct:    63 PGEWPWQASVRRQGAHICSGSLVADTWVLTAAHCFEKAAATELNSWSVVLGSLQREGLSP 122

Query:   495 --EEWGLKQLILHGAYTHPEGGYDMALLLLAQPVTLGASLRPLCLPYADHHLPDGERGWV 668
               || |+  |  |   ||| |     | |+||| || | |      |||||   |   |  |   |
Sbjct:   123 GAEEVGVAALQLPRAYNHYSQGSDLALLQLAHPTTH----TPLCLPQPAHRFPFGASCWA 178

Query:   669 LGRARPGAGI-SSLQTVPVTLLGPRACS----RLHAAPGGDGSPILPGMVCTSAVGELPS 833
             |  +  +    +|+ + + |+     |+     +||      +  |   |||+|     |  |
Sbjct:   179 TGWDQDTSDAPGTLRNLRLRLISRPTCNCIYNQLHQRHLSN--PARPGMLCG---GPQPG 233

Query:   834 CEANQPAADRGPGHSQEQENAGRQMALLPLSS 929
             +      ||      | +      |   ++  +|
Sbjct:   234 VQGPCQGDSGGPVLCLEPDGHWVQAGIISFAS 265

 Score = 125 (44.0 bits), Expect = 0.00067, P = 0.00067 (SEQ ID NO:119)
 Identities = 32/95 (33%), Positives = 47/95 (49%), Frame = +3

Query:    15 NPARPGMLCGGPQPGVQGPCQGDSGGPVLCLEPDGHWVQAGIISFASSCAQEDAPVLLTN 194
             +|  |||+|       |      |+| || |  |   |   | |  ||+ ||   +|       | + |
Sbjct:   474 SPILPGMVCTSAV-GELPSCEGLSGAP-LVHEVRGTWFLAGLHSFGDACQGPARPAVFTA 531

Query:   195 TAAHSSWLQARVQGAAFLAQSPETPEMSDEDSCVA 299
             |+  |+ + +      + |+ || || ++  ||+|
Sbjct:   532 LPAYEDWVSS-LDWQVYFAEEPE-PE-AEPGSCLA 563


>patp:Y90291 Human peptidase, HPEP-8 protein sequence - Homo sapiens, 267 aa.
(SEQ ID NO:74)

            Length = 267

Plus Strand HSPs:

 Score = 1028 (361.9 bits), Expect = 5.0e-103, P = 5.0e-103
 Identities = 189/189 (100%), Positives = 189/189 (100%), Frame = +3

Query:   273 MSDEDSCVACGSLRTAGPQAGAPSPWPWEARLMHQGQLACGGALVSEEAVLTAAHCFIGR 452
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:     1 MSDEDSCVACGSLRTAGPQAGAPSPWPWEARLMHQGQLACGGALVSEEAVLTAAHCFIGR 60

Query:   453 QAPEEWSVGLGTRPEEWGLKQLILHGAYTHPEGGYDMALLLLAQPVTLGASLRPLCLPYA 632
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    61 QAPEEWSVGLGTRPEEWGLKQLILHGAYTHPEGGYDMALLLLAQPVTLGASLRPLCLPYA 120

Query:   633 DHHLPDGERGWVLGRARPGAGISSLQTVPVTLLGPRACSRLHAAPGGDGSPILPGMVCTS 812
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   121 DHHLPDGERGWVLGRARPGAGISSLQTVPVTLLGPRACSRLHAAPGGDGSPILPGMVCTS 180

Query:   813 AVGELPSCE 839
             |||||||||
Sbjct:   181 AVGELPSCE 189
```

```
Score = 125 (44.0 bits), Expect = 0.00016, P = 0.00016 (SEQ ID NO:120)
Identities = 32/95 (33%), Positives = 47/95 (49%), Frame = +3

Query:    15 NPARPGMLCGGPQPGVQGPCQGDSGGPVLCLEPDGHWVQAGIISFASSCAQEDAPVLLTN 194
                +|  |||+|      |     |+| || | |  |  | |  ||+ ||  +|    | + |
Sbjct:   170 SPILPGMVCTSAV-GELPSCEGLSGAP-LVHEVRGTWFLAGLHSFGDACQGPARPAVFTA 227

Query:   195 TAAHSSWLQARVQGAAFLAQSPETPEMSDEDSCVA 299
                  |+  |+ + +     + |+ || || ++  ||+|
Sbjct:   228 LPAYEDWVSS-LDWQVYFAEEPE-PE-AEPGSCLA 259
```

Table 16. BLASTN identity search (versus the human SeqCalling database for the Peptidase-like protein of the invention.

```
>s3aq:132854740 Category D: 12 frag (12 non-5'sig-CG), 636 bp. (SEQ ID NO:75)
            Length = 636

 Minus Strand HSPs:

Score = 1423 (213.5 bits), Expect = 7.0e-59, P = 7.0e-59
Identities = 313/343 (91%), Positives = 313/343 (91%), Strand = Minus / Plus

Query:   754 AGCCGGCTGCAG-GCCCTAGGCCCCAGGAGGGTCACGGGCACTGTCTGGAGGGAGCTGAT 696
             ||| ||||||   | ||| |       ||| || ||   || ||| |     | |  |  |
Sbjct:   295 AGCTGGCTGCCCCGGCCT-GCAGGTTGGATGGACAGCAGCCCTGGCCCT-GTGCCCACCT 352

Query:   695 GCCTGCTCCTGGGCGGGCCCGTCCCAGAACCCAGCCACGCTCCCCATCAGGCAGGTGGTG 636
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   353 ACCTGCTCCTGGGCGGGCCCGTCCCAGAACCCAGCCACGCTCCCCATCAGGCAGGTGGTG 412

Query:   635 GTCAGCATAGGGCAGGCAGAGGGGCCGCAGGCTGGCTCCCAGTGTCACAGGCTGGGCCAG 576
             ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   413 GTCAGGATAGGGCAGGCAGAGGGGCCGCAGGCTGGCTCCCAGTGTCACAGGCTGGGCCAG 472

Query:   575 CAGCAGGAGGGCCATGTCGTAGCCCCCCTCAGGGTGGGTGTAGGCTCCATGCAGGATGAG 516
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   473 CAGCAGGAGGGCCATGTCGTAGCCCCCCTCAGGGTGGGTGTAGGCTCCATGCAGGATGAG 532

Query:   515 CTGCTTCAGGCCCCACTCCTCCGGTCTGGTCCCCAGCCCTACGCTCCATTCCTCTGGGGC 456
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   533 CTGCTTCAGGCCCCACTCCTCCGGTCTGGTCCCCAGCCCTACGCTCCATTCCTCTGGGGC 592

Query:   455 CTGGCGCCCAATGAAGCAGTGGGCAGCAGTTAGCACCGCCTCCT 412
             ||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   593 CTGGCGCCCAATGAAGCAGTGGGCAGCAGTTAGCACCGCCTCCT 636

Score = 757 (113.6 bits), Expect = 8.5e-29, P = 8.5e-29 (SEQ ID NO:121)
Identities = 165/179 (92%), Positives = 165/179 (92%), Strand = Minus / Plus

Query:   869 AGGTCCCCTGTCAGCAGCTGGTTGGTTGGCCTCACAGCTGGGCAGCTCACCCACAGCACT 810
             ||||    |||  |    ||||  || |  ||||||||||||||||||||||||||||||
Sbjct:   105 AGGTAAGGTGTGGGGGCCTGG--GGCTCACCTCACAGCTGGGCAGCTCACCCACAGCACT 162

Query:   809 GGTACACACCATCCCCGGCAGAATAGGGCTGCCATCACCCCCAGGAGCTGCATGCAGCCG 750
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   163 GGTACACACCATCCCCGGCAGAATAGGGCTGCCATCACCCCCAGGAGCTGCATGCAGCCG 222

Query:   749 GCTGCAGGCCCTAGGCCCCAGGAGGGTCACGGGCACTGTCTGGAGGGAGCTGATGCCTG 691
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   223 GCTGCAGGCCCTAGGCCCCAGGAGGGTCACGGGCACTGTCTGGAGGGAGCTGATGCCTG 281

>s3aq:134913963 Category E: 1 frag (1 non-CG EST), 415 bp. (SEQ ID NO:76)
```

```
          Length = 415

 Plus Strand HSPs:

Score = 297 (44.6 bits), Expect = 8.0e-07, P = 8.0e-07
Identities = 61/63 (96%), Positives = 61/63 (96%), Strand = Plus / Plus

Query:  1138 TTGTTTTGAAAATTTCTTTTTTTGGGGGGCAGCAGTTTTCCTTTTTTTAAACTTAAATAA 1197
             ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    10 TTGGTGTGAAAATTTCTTTTTTTGGGGGGCAGCAGTTTTCCTTTTTTTAAACTTAAATAA 69

Query:  1198 ATT 1200
             |||
Sbjct:    70 ATT 72
```

Table 17. ClustalW alignment of the protein of the invention.

```
CG50817-06  ------------------------------------------------------------
Y41704      MLLSSLVSLAGSVYLAWILFFVLYDFCIVCITTYAINVSLMWLSFRKVQEPQGKAKRHGN
Y90291      ------------------------------------------------------------

CG50817-06  ------------------------------------------------------------
Y41704      TVPGEWPWQASVRRQGAHICSGSLVADTWVLTAAHCFEKAAATELNSWSVVLGSLQREGL
Y90291      ------------------------------------------------------------

CG50817-06  ------------------------------------------------------------
Y41704      SPGAEEVGVAALQLPRAYNHYSQGSDLALLQLAHPTTHTPLCLPQPAHRFPFGASCWATG
Y90291      ------------------------------------------------------------

CG50817-06  --------------------------------------------MLCGGPQPGVQGPCQG
Y41704      WDQDTSDAPGTLRNLRLRLISRPTCNCIYNQLHQRHLSNPARPGMLCGGPQPGVQGPCQG
Y90291      ------------------------------------------------------------

CG50817-06  DSGGPVLCLEPDGHWVQAGIISFASSCAQEDAPVLLTNTAAHSSWLQARVQGAAFLAQSP
Y41704      DSGGPVLCLEPDGHWVQAGIISFASSCAQEDAPVLLTNTAAHSSWLQARVQGAAFLAQSP
Y90291      ------------------------------------------------------------

CG50817-06  ETPEMSDEDSCVACGSLRTAGPQAGAPSPWPWEARLMHQGQLACGGALVSEEAVLTAAHC
Y41704      ETPEMSDEDSCVACGSLRTAGPQAGAPSPWPWEARLMHQGQLACGGALVSEEAVLTAAHC
Y90291      ----MSDEDSCVACGSLRTAGPQAGAPSPWPWEARLMHQGQLACGGALVSEEAVLTAAHC

CG50817-06  FIGRQAPEEWSVGLGTRPEEWGLKQLILHGAYTHPEGGYDMALLLLAQPVTLGASLRPLC
Y41704      FIGRQAPEEWSVGLGTRPEEWGLKQLILHGAYTHPEGGYDMALLLLAQPVTLGASLRPLC
Y90291      FIGRQAPEEWSVGLGTRPEEWGLKQLILHGAYTHPEGGYDMALLLLAQPVTLGASLRPLC

CG50817-06  LPYADHHLPDGERGWVLGRARPGAGISSLQTVPVTLLGPRACSRLHAAPGGDGSPILPGM
Y41704      LPYPDHHLPDGERGWVLGRARPGAGISSLQTVPVTLLGPRACSRLHAAPGGDGSPILPGM
Y90291      LPYADHHLPDGERGWVLGRARPGAGISSLQTVPVTLLGPRACSRLHAAPGGDGSPILPGM

CG50817-06  VCTSAVGELPSCEANQPAADRGPG---------HSQEQENAGRQMALLPLSSFPCHV---
Y41704      VCTSAVGELPSCEGLSGAPLVHEVRGTWFLAGLHSFGDACQGPARPAVFTALPAYEDWVS
Y90291      VCTSAVGELPSCEGLSGAPLVHEVRGTWFLAGLHSFGDACQGPARPAVFTALPAYEDWVS

CG50817-06  --------------------------------
Y41704      SLDWQVVYFAEEPEPEAEPGSCLANISQPTSC
Y90291      SLDWQVVYFAEEPEPEAEPGSCLANISQPTSC
```

Information for the ClustalW proteins:

Accno	Common Name	Length

CG50817-06 **(SEQ ID NO:47)**	novel Peptidase-like protein	
Y41704 **(SEQ ID NO:122)**	Human PRO351 protein sequence.	571
Y90291 **(SEQ ID NO:123)**	Human peptidase, HPEP-8 protein sequence.	267

In the alignment shown above, black outlined amino acid residues indicate regions of conserved sequence (i.e., regions that may be required to preserve structural or functional properties); greyed amino acid residues can be mutated to a residue with comparable steric and/or chemical properties without altering protein structure or function (e.g. L to V, I, or M); non-highlighted amino acid residues can potentially be mutated to a much broader extent without altering structure or function.

Table 18. Psort, Signal P and hydropathy results for CG50817-06

```
                cytoplasm --- Certainty=0.4500(Affirmative) < succ>
     microbody (peroxisome) --- Certainty=0.3000(Affirmative) < succ>
           lysosome (lumen) --- Certainty=0.2334(Affirmative) < succ>
 mitochondrial matrix space --- Certainty=0.1000(Affirmative) < succ>


Is the sequence a signal peptide?
# Measure  Position  Value  Cutoff  Conclusion
  max. C    45       0.253   0.37   NO
  max. Y    17       0.064   0.34   NO
  max. S    68       0.536   0.88   NO
  mean S     1-16    0.130   0.48   NO
```

SECP 14

A SECP14 nucleic acid and polypeptide according to the invention includes the nucleic acid sequence (SEQ ID NO:48) and encoded polypeptide sequence (SEQ ID NO:49) of clone CG50817-06 directed toward novel serine protease-like proteins and nucleic acids encoding them. Figure 19 illustrates the nucleic acid sequence and amino acid sequences respectively. This clone includes a nucleotide sequence (SEQ ID NO:48) of 1214 bp. The nucleotide sequence includes an open reading frame (ORF) beginning with an ATG initiation codon at nucleotides 31-33 and ending at nucleotides 1186-1188. Putative untranslated regions, if any, are found upstream from the initiation codon and downstream from the termination codon. The encoded protein having 385 amino acid residues is presented using the one-letter code in Figure 19. The protein encoded by clone CG51099-03 is predicted by the PSORT program to the outside of the membrane with a certainty of 0.5804, and appears to be a signal protein (see Table 22 below).

The serine protease tryptase (ECNr. 3.4. 21.59), which is almost exclusively expressed in mast cells, is released by mast cell degranulation in an enzymatically active form together with other mediators, e.g. histamine, into the extracellular space and the circulation. The capability of the enzyme to directly stimulate several cell types as well as to cleave polypeptide hormones and to activate pro-enzymes suggests a role for tryptase in inflammatory and tissue-remodeling processes. Therefore, in the skin, a role of tryptase is suggested not only in mastocytosis and immediate type hypersensitivity reactions, but also in other inflammatory diseases, degenerative or neoplastic conditions as well as in wound healing, where an accumulation and/or activation of mast cells is found. Extracellular tryptase may be superior to histamine as a parameter for the onset and course of immediate type reactions and as an indicator for the activation of mast cells in other conditions. Its absence during histamine-liberating reactions may suggest basophil activation. In addition, tryptase has been shown to be a sensitive and specific marker for the localization of mast cells in tissues (Ludolf-Hauser et al., 1999, Hautarzt 50:556-61).

Tryptases are stored in abundance in the secretory granules of mouse (McNeil et al, 1992, Proc. Natl. Acad. Sci. U. S. A. 89, 11174-11178; Johnson, D. A., and Barton, G., 1992, Protein Sci. 1, 370-377), and human (Vanderslice et al., 1990, Proc. Natl. Acad. Sci. U. S. A. 87, 3811-3815) mast cells (MCs). In humans, the four homologous tryptases (designated tryptases I, II/ , III, and) that have been cloned reside at a complex on chromosome 16 (Pallaoro et al., 1999, J. Biol. Chem. 274, 3355-3362). Although only two tryptases (designated mouse MC protease (mMCP) 6 and mMCP-7) have been identified so far in the mouse, their genes reside ~1.2 centimorgans away from each other on the syntenic region of mouse chromosome 17 (Gurish et al., 1994, Mammal. Genome 5, 656-657). Despite the chromosomal clustering of their genes, these mouse tryptases are differentially regulated in vivo (Reynolds et al., 1990, Proc. Natl. Acad. Sci. U. S. A. 87, 3230-3234) and in vitro (Reynolds et al., 1991, J. Biol. Chem. 266, 3847-3853; McNeil et al, 1992, Proc. Natl. Acad. Sci. U. S. A. 89, 11174-11178) at the levels of gene transcription (Morri et al., 1996, Blood 88, 2488-2494) and mRNA stability.

All known mouse and human tryptases in this family are initially translated as zymogens. They possess an ~20-residue hydrophobic signal peptide which is presumed to be removed in the endoplasmic reticulum immediately after the translated zymogen is translocated into the lumen. They also possess an ~10-residue propeptide preceding the mature portion of the enzyme which consists of ~245 amino acids. Although tryptases undergo variable N-linked glycosylation during

their biosynthesis (Ghidyal et al., 1994, J. Immunol. 153, 2624-2630), the current members of the family appear to be targeted to the secretory granule by a serglycin proteoglycan-dependent mechanism (Ghidyal et al., 1996, J. Exp. Med. 184, 1061-1073) rather than by a Man-PO4-dependent mechanism as are classical lysosomal enzymes.

Recently, Wong et al. (1999, J Biol Chem 274, 30784-30793) described a novel mouse gene, and its human ortholog, which encode an unusual transmembrane tryptase (TMT). Comparative structural studies indicated that the putative transmembrane tryptase (TMT) possesses a unique substrate-binding cleft. As assessed by RNA blot analyses, mTMT is expressed in mice in both strain- and tissue-dependent manners. Thus, different transcriptional and/or post-transcriptional mechanisms are used to control the expression of mTMT in vivo. Analysis of the corresponding tryptase locus in the human genome resulted in the isolation and characterization of the hTMT gene. The hTMT transcript is expressed in numerous tissues and is also translated. Analysis of the tryptase family of genes in mice and humans now indicates that a primordial serine protease gene duplicated early and often during the evolution of mammals to generate a panel of homologous tryptases in each species that differ in their tissue expression, substrate specificities, and physical properties.

Similarities

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 1213 of 1213 bases (100%) identical to a gb:GENBANK-ID:AX079882|acc:AX079882.1 mRNA from Homo sapiens (Sequence 13 from Patent WO0105971) (See Table 19). The full amino acid sequence of the protein of the invention was found to have 385 of 385 amino acid residues (100%) identical to, and 385 of 385 amino acid residues (100%) similar to, the 385 amino acid residue ptnr:SPTREMBL-ACC:Q9UI38 protein from Homo sapiens (Human) (TESTES-SPECIFIC PROTEIN TSP50)(See Table 20).

A multiple sequence alignment is given in Table 21, with the protein of the invention being shown on the first line in a ClustalW analysis comparing the protein of the invention with related protein sequences.

The presence of identifiable domains in the protein disclosed herein was determined by searches versus domain databases such as Pfam, PROSITE, ProDom, Blocks or Prints and then identified by the Interpro domain accession number. Significant domains are summarized below:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
trypsin	1/2	118	297	6	199	104.4	2.6e-32
trypsin	2/2	313	353	215	259	35.9	1.6e-10

The catalytic activity of the serine proteases from the trypsin family is provided by a charge relay system involving an aspartic acid residue hydrogen-bonded to a histidine, which itself is hydrogen-bonded to a serine. The sequences in the vicinity of the active site serine and histidine residues are well conserved in this family of proteases (Sprang et al., 1987, Science 237:905-909). A partial list of proteases known to belong to the trypsin family is shown below.

- Acrosin.
- Blood coagulation factors VII, IX, X, XI and XII, thrombin, plasminogen, and protein C.
- Cathepsin G.
- Chymotrypsins.
- Complement components C1r, C1s, C2, and complement factors B, D and I.
- Complement-activating component of RA-reactive factor.
- Cytotoxic cell proteases (granzymes A to H).
- Duodenase I.
- Elastases 1, 2, 3A, 3B (protease E), leukocyte (medullasin).
- Enterokinase (EC 3.4.21.9) (enteropeptidase).
- Hepatocyte growth factor activator.
- Hepsin.
- Glandular (tissue) kallikreins (including EGF-binding protein types A, B, and C, NGF-gamma chain, gamma-renin, prostate specific antigen (PSA) and tonin).
- Plasma kallikrein.
- Mast cell proteases (MCP) 1 (chymase) to 8.
- Myeloblastin (proteinase 3) (Wegener's autoantigen).
- Plasminogen activators (urokinase-type, and tissue-type).
- Trypsins I, II, III, and IV.
- Tryptases.
- Snake venom proteases such as ancrod, batroxobin, cerastobin, flavoxobin, and protein C activator.
- Collagenase from common cattle grub and collagenolytic protease from Atlantic sand fiddler crab.
- Apolipoprotein(a).
- Blood fluke cercarial protease.
- Drosophila trypsin like proteases: alpha, easter, snake-locus.
- Drosophila protease stübble (gene sb).
- Major mite fecal allergen Der p III.

All the above proteins belong to family S1 in the classification of peptidases.

This indicates that the sequence of the invention has properties similar to those of other proteins known to contain this/these domain(s) and similar to the properties of these domains.

Chromosomal information:

The Serine Protease-like gene disclosed in this invention maps to chromosome 3. This assignment was made using mapping information associated with genomic clones, public genes and ESTs sharing sequence identity with the disclosed sequence and CuraGen Corporation's Electronic Northern bioinformatic tool.

Tissue expression

The Serine Protease-like gene disclosed in this invention is expressed in at least the following tissues: adipose, adrenal gland, thyroid, brain, heart, skeletal muscle, bone marrow, colon, bladder, liver, lung, mammary gland, placenta, testis. Expression information was derived from the tissue sources of the sequences that were included in the derivation of the sequence of CuraGen Acc. No. CG51099-03.The sequence is predicted to be expressed in the following tissues because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:AX079882|acc:AX079882.1) a closely related Sequence 13 from Patent WO0105971 homolog in species Homo sapiens: testis.

Cellular Localization and Sorting

The PSORT, SignalP and hydropathy profile for the Serine Protease-like protein are shown in Table 22. The results predict that this sequence has a signal peptide and is likely to be localized extracellularly with a certainty of 0.5804. The signal peptide is predicted by SignalP to be cleaved at amino acid 39 and 40: CWG-AG.

Functional Variants and Homologs

The novel nucleic acid of the invention encoding a Serine Protease-like protein includes the nucleic acid whose sequence is provided in Figure 19, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Figure 19 while still encoding a protein that maintains its Serine Protease-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to the sequence of CuraGen Acc. No. CG51099-03, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such

modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 0% of the bases may be so changed.

The novel protein of the invention includes the Serine Protease-like protein whose sequence is provided in Figure 19. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Figure 19 while still encoding a protein that maintains its Serine Protease-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 0% of the amino acid residues may be so changed.

Antibodies

The invention further encompasses antibodies and antibody fragments, such as Fab, $(Fab)_2$ or single chain FV constructs, that bind immunospecifically to any of the proteins of the invention. Also encompassed within the invention are peptides and polypeptides comprising sequences having high binding affinity for any of the proteins of the invention, including such peptides and polypeptides that are fused to any carrier particle (or biologically expressed on the surface of a carrier) such as a bacteriophage particle.

Uses of the Compositions of the Invention

The protein similarity information, expression pattern, cellular localization, and map location for the protein and nucleic acid disclosed herein suggest that this Serine Protease-like protein may have important structural and/or physiological functions characteristic of the Trypsin family. Therefore, the nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications and as a research tool. These include serving as a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed. These also include potential therapeutic applications such as the following: (i) a protein therapeutic, (ii) a small molecule drug target, (iii) an antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody),

(iv) a nucleic acid useful in gene therapy (gene delivery/gene ablation), (v) an agent promoting tissue regeneration *in vitro* and *in vivo*, and (vi) a biological defense weapon.

The nucleic acids and proteins of the invention have applications in the diagnosis and/or treatment of various diseases and disorders. For example, the compositions of the present invention will have efficacy for the treatment of patients suffering from: adrenoleukodystrophy , congenital adrenal hyperplasia, hyperthyroidism, hypothyroidism, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalceimia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neurodegeneration, cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, scleroderma, obesity, transplantation, muscular dystrophy, myasthenia gravis, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, graft versus host disease, cirrhosis, systemic lupus erythematosus, asthma, emphysema, ARDS, fertility, cancer, as well as other diseases, disorders and conditions.

These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in diagnostic and/or therapeutic methods.

Table 19. BLASTN search using CuraGen Acc. No. CG51099-03.

```
>gb:GENBANK-ID:AX079882|acc:AX079882.1 Sequence 13 from Patent WO0105971 - Homo
sapiens, 1359 bp. (SEQ ID NO:77)
            Length = 1359

  Plus Strand HSPs:

 Score = 6065 (910.0 bits), Expect = 4.8e-268, P = 4.8e-268
 Identities = 1213/1213 (100%), Positives = 1213/1213 (100%), Strand = Plus / Plus

Query:      1 CGGAGAGACGCAGTCGGCTGCCACCCCGGGATGGGTCGCTGGTGCCAGACCGTCGCGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:     15 CGGAGAGACGCAGTCGGCTGCCACCCCGGGATGGGTCGCTGGTGCCAGACCGTCGCGCGC 74

Query:     61 GGGCAGCGCCCCCGGACGTCTGCCCCCTCCCGCGCCGGTGCCCTGCTGCTGCTGCTTCTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:     75 GGGCAGCGCCCCCGGACGTCTGCCCCCTCCCGCGCCGGTGCCCTGCTGCTGCTGCTTCTG 134

Query:    121 TTGCTGAGGTCTGCAGGTTGCTGGGGCGCAGGGGAAGCCCCGGGGGCGCTGTCCACTGCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    135 TTGCTGAGGTCTGCAGGTTGCTGGGGCGCAGGGGAAGCCCCGGGGGCGCTGTCCACTGCT 194

Query:    181 GATCCCGCCGACCAGAGCGTCCAGTGTGTCCCCAAGGCCACCTGTCCTTCCAGCCGGCCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

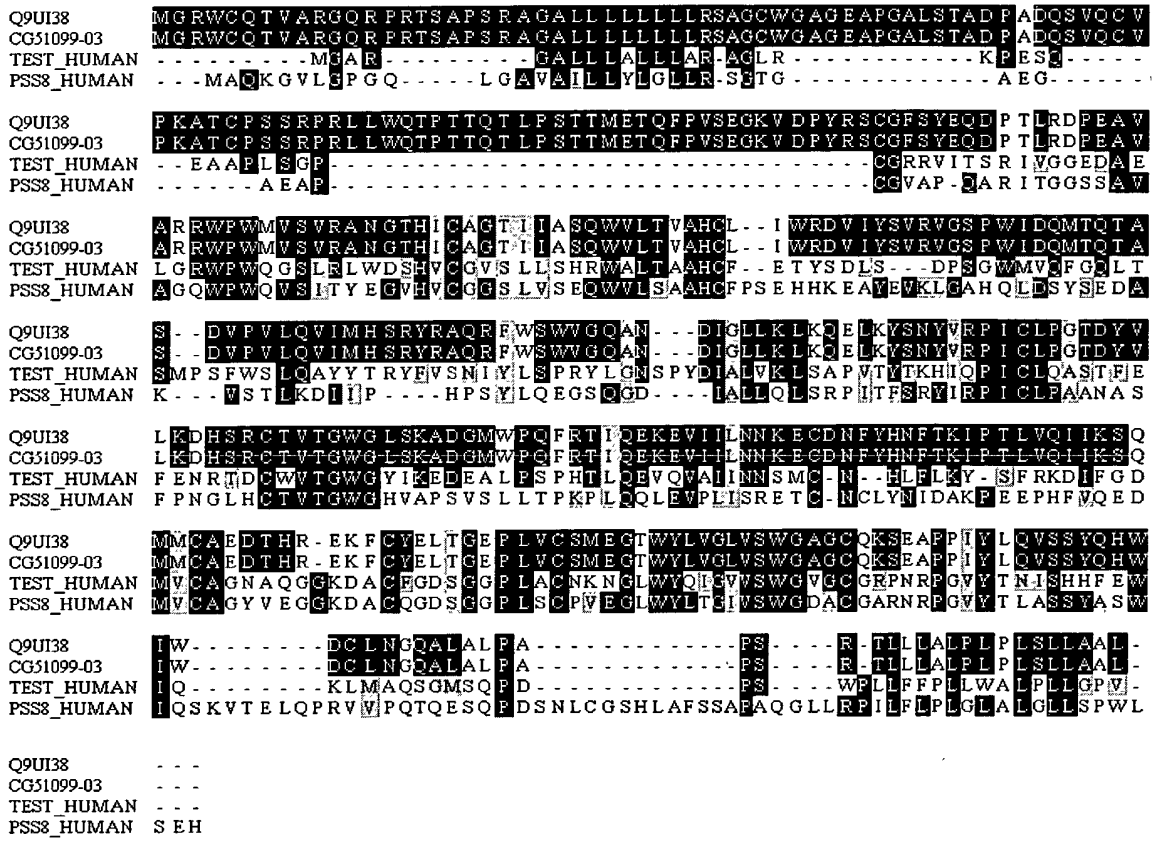
```
Sbjct:  195 GATCCCGCCGACCAGAGCGTCCAGTGTGTCCCCAAGGCCACCTGTCCTTCCAGCCGGCCT 254

Query:  241 CGCCTTCTCTGGCAGACCCCGACCACCCAGACACTGCCCTCGACCACCATGGAGACCCAA 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  255 CGCCTTCTCTGGCAGACCCCGACCACCCAGACACTGCCCTCGACCACCATGGAGACCCAA 314

Query:  301 TTCCCAGTTTCTGAAGGCAAAGTCGACCCATACCGCTCCTGTGGCTTTTCCTACGAGCAG 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  315 TTCCCAGTTTCTGAAGGCAAAGTCGACCCATACCGCTCCTGTGGCTTTTCCTACGAGCAG 374

Query:  361 GACCCCACCCTCAGGGACCCAGAAGCCGTGGCTCGGCGGTGGCCCTGGATGGTCAGCGTG 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  375 GACCCCACCCTCAGGGACCCAGAAGCCGTGGCTCGGCGGTGGCCCTGGATGGTCAGCGTG 434

Query:  421 CGGGCCAATGGCACACACATCTGTGCCGGCACCATCATTGCCTCCCAGTGGGTGCTGACT 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  435 CGGGCCAATGGCACACACATCTGTGCCGGCACCATCATTGCCTCCCAGTGGGTGCTGACT 494

Query:  481 GTGGCCCACTGCCTGATCTGGCGTGATGTTATCTACTCAGTGAGGGTGGGGAGTCCGTGG 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  495 GTGGCCCACTGCCTGATCTGGCGTGATGTTATCTACTCAGTGAGGGTGGGGAGTCCGTGG 554

Query:  541 ATTGACCAGATGACGCAGACCGCCTCCGATGTCCCGGTGCTCCAGGTCATCATGCATAGC 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  555 ATTGACCAGATGACGCAGACCGCCTCCGATGTCCCGGTGCTCCAGGTCATCATGCATAGC 614

Query:  601 AGGTACCGGGCCCAGCGGTTCTGGTCCTGGGTGGGCCAGGCCAACGACATCGGCCTCCTC 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  615 AGGTACCGGGCCCAGCGGTTCTGGTCCTGGGTGGGCCAGGCCAACGACATCGGCCTCCTC 674

Query:  661 AAGCTCAAGCAGGAACTCAAGTACAGCAATTACGTGCGGCCCATCTGCCTGCCTGGCACG 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  675 AAGCTCAAGCAGGAACTCAAGTACAGCAATTACGTGCGGCCCATCTGCCTGCCTGGCACG 734

Query:  721 GACTATGTGTTGAAGGACCATTCCCGCTGCACTGTGACGGGCTGGGGACTTTCCAAGGCT 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  735 GACTATGTGTTGAAGGACCATTCCCGCTGCACTGTGACGGGCTGGGGACTTTCCAAGGCT 794

Query:  781 GACGGCATGTGGCCTCAGTTCCGGACCATTCAGGAGAAGGAAGTCATCATCCTGAACAAC 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  795 GACGGCATGTGGCCTCAGTTCCGGACCATTCAGGAGAAGGAAGTCATCATCCTGAACAAC 854

Query:  841 AAAGAGTGTGACAATTTCTACCACAACTTCACCAAAATCCCCACTCTGGTTCAGATCATC 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  855 AAAGAGTGTGACAATTTCTACCACAACTTCACCAAAATCCCCACTCTGGTTCAGATCATC 914

Query:  901 AAGTCCCAGATGATGTGTGCGGAGGACACCCACAGGGAGAAGTTCTGCTATGAGCTAACT 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  915 AAGTCCCAGATGATGTGTGCGGAGGACACCCACAGGGAGAAGTTCTGCTATGAGCTAACT 974

Query:  961 GGAGAGCCCTTGGTCTGCTCCATGGAGGGCACGTGGTACCTGGTGGGATTGGTGAGCTGG 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  975 GGAGAGCCCTTGGTCTGCTCCATGGAGGGCACGTGGTACCTGGTGGGATTGGTGAGCTGG 1034

Query: 1021 GGTGCAGGCTGCCAGAAGAGCGAGGCCCCACCCATCTACCTACAGGTCTCCTCCTACCAA 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1035 GGTGCAGGCTGCCAGAAGAGCGAGGCCCCACCCATCTACCTACAGGTCTCCTCCTACCAA 1094

Query: 1081 CACTGGATCTGGGACTGCCTCAACGGGCAGGCCCTGGCCCTGCCAGCCCCATCCAGGACC 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1095 CACTGGATCTGGGACTGCCTCAACGGGCAGGCCCTGGCCCTGCCAGCCCCATCCAGGACC 1154

Query: 1141 CTGCTCCTGGCACTCCCACTGCCCCTCAGCCTCCTTGCTGCCCTCTGACTCTGTGTGCCC 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1155 CTGCTCCTGGCACTCCCACTGCCCCTCAGCCTCCTTGCTGCCCTCTGACTCTGTGTGCCC 1214

Query: 1201 TCCCTCACTTGTG 1213
            |||||||||||||
```

```
Sbjct:  1215 TCCCTCACTTGTG 1227
```

Table 20. BLASTP search using the protein of CuraGen Acc. No. CG51099-03.

```
>ptnr:SPTREMBL-ACC:Q9UI38 TESTES-SPECIFIC PROTEIN TSP50 - Homo sapiens (Human),
            385 aa. (SEQ ID NO:78)
            Length = 385

 Score = 2090 (735.7 bits), Expect = 4.5e-216, P = 4.5e-216
 Identities = 385/385 (100%), Positives = 385/385 (100%)

Query:      1 MGRWCQTVARGQRPRTSAPSRAGALLLLLLLLRSAGCWGAGEAPGALSTADPADQSVQCV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:      1 MGRWCQTVARGQRPRTSAPSRAGALLLLLLLLRSAGCWGAGEAPGALSTADPADQSVQCV 60

Query:     61 PKATCPSSRPRLLWQTPTTQTLPSTTMETQFPVSEGKVDPYRSCGFSYEQDPTLRDPEAV 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:     61 PKATCPSSRPRLLWQTPTTQTLPSTTMETQFPVSEGKVDPYRSCGFSYEQDPTLRDPEAV 120

Query:    121 ARRWPWMVSVRANGTHICAGTIIASQWVLTVAHCLIWRDVIYSVRVGSPWIDQMTQTASD 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    121 ARRWPWMVSVRANGTHICAGTIIASQWVLTVAHCLIWRDVIYSVRVGSPWIDQMTQTASD 180

Query:    181 VPVLQVIMHSRYRAQRFWSWVGQANDIGLLKLKQELKYSNYVRPICLPGTDYVLKDHSRC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    181 VPVLQVIMHSRYRAQRFWSWVGQANDIGLLKLKQELKYSNYVRPICLPGTDYVLKDHSRC 240

Query:    241 TVTGWGLSKADGMWPQFRTIQEKEVIILNNKECDNFYHNFTKIPTLVQIIKSQMMCAEDT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    241 TVTGWGLSKADGMWPQFRTIQEKEVIILNNKECDNFYHNFTKIPTLVQIIKSQMMCAEDT 300

Query:    301 HREKFCYELTGEPLVCSMEGTWYLVGLVSWGAGCQKSEAPPIYLQVSSYQHWIWDCLNGQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    301 HREKFCYELTGEPLVCSMEGTWYLVGLVSWGAGCQKSEAPPIYLQVSSYQHWIWDCLNGQ 360

Query:    361 ALALPAPSRTLLLALPLPLSLLAAL 385
              |||||||||||||||||||||||||
Sbjct:    361 ALALPAPSRTLLLALPLPLSLLAAL 385
```

Table 21. ClustalW alignment of CG51099-03 protein with related proteins.



Information for the ClustalW proteins:

Accno	Common Name	Length
CG51099-03 **(SEQ ID NO:49)**	novel Serine Protease-like protein	
TEST_HUMAN **(SEQ ID NO:124)**	TESTISIN PRECURSOR (EC 3.4.21.-) (EOSINOPHIL SERINE PROTEASE 1) (ESP- DE 1).	314
PSS8_HUMAN **(SEQ ID NO:125)**	PROSTASIN PRECURSOR (EC 3.4.21.-).	343
Q9UI38 **(SEQ ID NO:78)**	TESTES-SPECIFIC PROTEIN TSP50.	385

In the alignment shown above, black outlined amino acid residues indicate residues identically conserved between sequences (i.e., residues that may be required to preserve structural or functional properties); amino acid residues with a gray background are similar to one another between sequences, possessing comparable physical and/or chemical properties

without altering protein structure or function (e.g. the group L, V, I, and M may be considered similar); and amino acid residues with a white background are neither conserved nor similar between sequences.

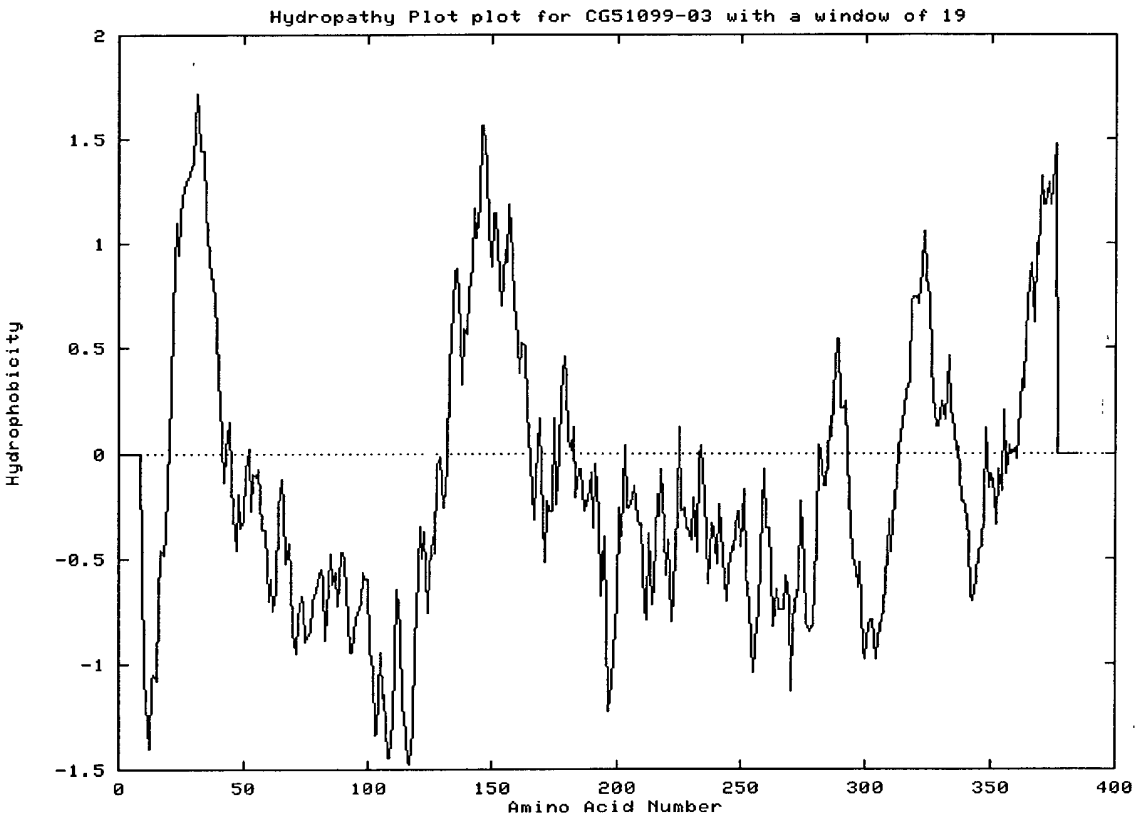
Table 22. PSORT, SignalP and hydropathy results for CuraGen Acc. No. CG51099-03.

```
outside --- Certainty=0.5804(Affirmative) < succ>
            lysosome (lumen) --- Certainty=0.5144(Affirmative) < succ>
         microbody (peroxisome) --- Certainty=0.1203(Affirmative) < succ>
 endoplasmic reticulum (membrane) --- Certainty=0.1000(Affirmative) < succ>
```

Is the sequence a signal peptide?

# Measure	Position	Value	Cutoff	Conclusion
max. C	40	0.888	0.37	YES
max. Y	40	0.848	0.34	YES
max. S	30	0.975	0.88	YES
mean S	1-39	0.708	0.48	YES

Most likely cleavage site between pos. 39 and 40: CWG-AG



SECP 15

A SECP15 nucleic acid and polypeptide according to the invention includes the nucleic acid sequence (SEQ ID NO:50) and encoded polypeptide sequence (SEQ ID NO:51) of clone

CG57051-04 directed toward novel Angiopoietin-like proteins and nucleic acids encoding them. Figure 20 illustrates the nucleic acid sequence and amino acid sequences respectively. This clone includes a nucleotide sequence (SEQ ID NO:50) of 937 bp. The nucleotide sequence includes an open reading frame (ORF) beginning with an ATG initiation codon at nucleotides 155-157 and ending with a TAG stop codon at nucleotides 881-883. Putative untranslated regions, if any, are found upstream from the initiation codon and downstream from the termination codon. The encoded protein having 242 amino acid residues is presented using the one-letter code in Figure 20. The protein encoded by clone CG57051-04 is predicted by the PSORT program to be located at the endoplasmic reticulum with a certainty of 0.8200, and appears to be a signal protein (see Table 27 below).

PPARG ANGIOPOIETIN-RELATED PROTEIN – PGAR:

Background

The peroxisome proliferator-activated receptors (PPARs) are members of the nuclear hormone receptor subfamily of transcription factors. PPARs form heterodimers with retinoid X receptors (RXRs) and these heterodimers regulate transcription of various genes. There are 3 known subtypes of PPARs, PPAR-alpha (170998), PPAR-delta (600409), and PPAR-gamma. PPAR-gamma is believed to be involved in adipocyte differentiation. Tontonoz et al. (1994) found 2 isoforms of PPAR-gamma in mouse, gamma-1 and gamma-2, resulting from the use of different initiator methionines.

Elbrecht et al. (1996) cloned cDNAs of PPAR-gamma-1 and PPAR-gamma-2 from human fat cell cDNA by PCR using primers based on the mouse sequence and on a previously published human cDNA sequence (Greene et al., 1995). They found that the human PPAR-gamma-1 and PPAR-gamma-2 genes have identical sequences except that PPAR-gamma-2 contains an additional 84 nucleotides at its 5-prime end. The sequences obtained by Elbrecht et al. (1996) differed at 3 sites from the previously published human PPAR-gamma-1 sequence of Greene et al. (1995). By Northern blot analysis, Elbrecht et al. (1996) found that human PPAR-

gamma is expressed at high levels in adipocytes and at a much lower level in bone marrow, spleen, testis, brain, skeletal muscle, and liver.

The thiazolidinediones are synthetic compounds that can normalize elevated plasma glucose levels in obese, diabetic rodents and may be efficacious therapeutic agents for the treatment of noninsulin-dependent diabetes mellitus. Lehmann et al. (1995) identified the thiazolidinediones as high affinity ligands for mouse PPAR-gamma receptors. Elbrecht et al. (1996) confirmed that human PPAR-gamma-1 and PPAR-gamma-2 have similar activity and determined that 3 different thiazolidinedione compounds are agonists of PPAR-gamma-1 and PPAR-gamma-2. Elbrecht et al. (1996) speculated that the antidiabetic activity of the thiazolidinediones in humans is mediated through the activation of PPAR-gamma-1 and PPAR-gamma-2.

The nuclear receptor PPARG/RXRA heterodimer regulates glucose and lipid homeostasis and is the target for the antidiabetic drugs GI262570 and the thiazolidinediones. Gampe et al. (2000) reported the crystal structures of the PPARG and RXRA ligand-binding domains complexed with the RXRA ligand 9-cis-retinoic acid, the PPARG agonist GI262570, and coactivator peptides. The structures provided a molecular understanding of the ability of RXRs to heterodimerize with many nuclear receptors and of the permissive activation of the PPARG/RXRA heterodimer by 9-cis-retinoic acid.

Mueller et al. (1998) showed that PPAR-gamma is expressed at significant levels in human primary and metastatic breast adenocarcinomas. Ligand activation of this receptor in cultured breast cancer cells caused extensive lipid accumulation, changes in breast epithelial gene expression associated with a more differentiated, less malignant state, and a reduction in growth rate and clonogenic capacity of the cells. Inhibition of MAP kinase, a powerful negative regulator of PPAR-gamma, improves the thiazolidinedione ligand sensitivity of nonresponsive cells. These data suggested that the PPAR-gamma transcriptional pathway can induce terminal differentiation of malignant breast epithelial cells.

Tontonoz et al. (1994) identified a novel adipocyte-specific transcription factor, which they termed ARF6, and showed that it is a heterodimeric complex of RXRA and PPARG. (This ARF6 is not to be confused with ADP-ribosylation factor 6 (600464), with is also symbolized ARF6.) Tontonoz et al. (1995) demonstrated that PPAR-gamma-2 regulates adipocyte expression of the phosphoenolpyruvate carboxykinase gene (PCK1, 261680; PCK2, 261650).

The formation of foam cells from macrophages in the arterial wall is characterized by dramatic changes in lipid metabolism, including increased expression of scavenger receptors and the uptake of oxidized low density lipoprotein (oxLDL). Tontonoz et al. (1998) demonstrated that the nuclear receptor PPAR-gamma is induced in human monocytes following exposure to oxLDL and is expressed at high levels in the foam cells of atherosclerotic lesions. Ligand activation of the PPAR-gamma:RXR-alpha heterodimer in myelomonocytic cell lines induced changes characteristic of monocytic differentiation and promoted uptake of oxLDL through transcriptional induction of the scavenger receptor CD36. These results revealed a novel signaling pathway controlling differentiation and lipid metabolism in monocytic cells. Tontonoz et al. (1998) suggested that endogenous PPAR-gamma ligands may be important regulators of gene expression during atherogenesis.

Nagy et al. (1998) demonstrated that oxLDL activates PPAR-gamma-dependent transcription through a signaling pathway involving scavenger receptor-mediated particle uptake. Moreover, they identified 2 of the major oxidized linoleic acid metabolite components of oxLDL, 9-HODE and 13-HODE, as endogenous activators and ligands of PPAR-gamma. The authors found that the biologic effects of oxLDL are coordinated by 2 sets of receptors, one on the cell surface, which binds and internalizes the particle, and one in the nucleus, which is transcriptionally activated by its component lipids. Nagy et al. (1998) suggested that PPAR-gamma may be a key regulator of foam cell gene expression.

Chawla et al. (2001) provided evidence that in addition to lipid uptake, PPARG regulates a pathway of cholesterol efflux. PPARG induces ABCA1 (600046) expression and cholesterol removal from macrophages through a transcriptional cascade mediated by the nuclear receptor LXRA (NR1H3; 602423). Ligand activation of PPARG leads to primary induction of LXRA and to coupled induction of ABCA1. Transplantation of PPAR null bone marrow into Ldlr -/- mice resulted in a significant increase in atherosclerosis, consistent with the hypothesis that regulation of LXRA and ABCA1 expression is protective in vivo. Chawla et al. (2001) proposed that PPARG coordinates a complex physiologic response to oxLDL that involves particle uptake, processing, and cholesterol removal through ABCA1.

Fajas et al. (1997) used competitive RT-PCR to distinguish relative PPARG1 and PPARG2 mRNA levels in tissues. They determined that PPARG2 is much less abundant than PPARG1. The highest levels of PPARG are found in adipose tissue and large intestine, with

intermediate levels in kidney, liver, and small intestine, and barely detectable levels in muscle. Western blot analysis showed that PPARG is expressed as a 60-kD protein. EMSA analysis indicated that PPARG2 binds to and transactivates through a peroxisome proliferator response element. The PPARG gene contains 9 exons and spans more than 100 kb. Through alternative transcription start sites and alternate splicing, the mRNAs differ at their 5-prime ends, with PPARG1 being encoded by 8 and PPARG2 by 7 exons. PPARG1 uses exons A1 and A2, whereas PPARG2 uses exon B; both use exons 1 through 6.

Martin et al. (1998) reported that there are 3 PPARG isoforms which differ at their 5-prime ends, each under the control of its own promoter. PPARG1 and PPARG3, however, give rise to the same protein, encoded by exons 1 through 6, because neither the A1 nor the A2 exon are translated. By RNase protection analysis, Ricote et al. (1998) showed that in phorbol ester-stimulated macrophage cell lines, a probe to PPARG1 protected a 218-nucleotide fragment of PPARG1, but only a 174-nucleotide fragment of PPARG3. A PPARG2 probe protected a common 104-nucleotide fragment of both PPARG1 and PPARG3. PPARG2 itself was not expressed in the stimulated macrophages. PPARG1 and PPARG2 promoters are primarily used in adipose tissue. The authors speculated that other inducing factors, such as cytokines MCSF (120420) or GMCSF (138960), or oxidized LDL (see OLR1, 602601), might differentially regulate expression of the 3 isoforms.

Lowell (1999) reviewed the role of PPARG in adipogenesis.

Kersten et al. (2000) reviewed the roles of PPARs in health and disease.

Tong et al. (2000) showed that murine GATA2 (137295) and GATA3 (131320) are specifically expressed in white adipocyte precursors and that their downregulation sets the stage for terminal differentiation. Constitutive GATA2 and GATA3 expression suppressed adipocyte differentiation and trapped cells at the preadipocyte stage. This effect was mediated, at least in part, through the direct suppression of PPARG.

Mueller et al. (2000) showed that PPAR-gamma is expressed in human prostate adenocarcinomas and cell lines derived from these tumors. Activation of this receptor with specific ligands exerts an inhibitory effect on the growth of prostate cancer (176807) cell lines. They showed that prostate cancer and cell lines do not have intragenic mutations in the PPARG gene, although 40% of the informative tumors have hemizygous deletions of this gene. They

conducted a phase II clinical study in patients with advanced prostate cancer using troglitazone (Rezulin), a PPAR-gamma ligand used for the treatment of type II diabetes. Oral treatment was administered to 41 men with histologically confirmed prostate cancer and no symptomatic metastatic disease. An unexpectedly high incidence of prolonged stabilization of prostate-specific antigen (KLK3; 176820) was seen in patients treated with troglitazone. In addition, 1 patient had a dramatic decrease in serum prostate-specific antigen to nearly undetectable levels. The findings suggested that PPAR-gamma may serve as a biologic modifier in human prostate cancer and that its therapeutic potential should be further studied.

By somatic cell hybridization and linkage analysis, Greene et al. (1995) mapped the human PPARG gene to 3p25. Beamer et al. (1997) mapped the gene to 3p25 by fluorescence in situ hybridization.

Meirhaeghe et al. (1998) detected a polymorphism corresponding to a silent C-to-T substitution in exon 6 of the PPARG gene (601487.0009).

Since PPARG is a transcription factor that has a key role in adipocyte differentiation, Ristow et al. (1998) investigated whether mutations of the gene encoding this factor predispose people to obesity. They studied 358 unrelated German subjects, including 121 obese subjects, looking for mutations in the PPARG2 gene at or near a site of serine phosphorylation at position 114 that negatively regulates transcriptional activity of the protein. Four of the 121 obese subjects had a missense mutation in the PPARG2 gene that resulted in conversion of proline to glutamine at position 115 (601487.0001), as compared with none of the 237 subjects of normal weight. All the subjects with the mutant allele were markedly obese. Overexpression of the mutant gene in murine fibroblasts led to the production of a protein in which the phosphorylation of serine at position 114 was defective, as well as accelerated differentiation of the cells into adipocytes and greater cellular accumulation of triglyceride than with the wildtype PPAR-gamma-2. These effects were similar to those of an in vitro mutation created directly at the ser114 phosphorylation site.

PPARG1 and PPARG2 have ligand-dependent and -independent activation domains. PPARG2 has an additional 28 amino acids at the amino terminus that render its ligand-independent activation domain 5- to 10-fold more effective than that of PPARG1. Insulin stimulates the ligand-independent activation of PPARG1 and PPARG2; however, obesity and nutritional factors influence only the expression of PPARG2 in human adipocytes. Deeb et al.

(1998) reported that a relatively common pro12-to-ala substitution in PPARG2 (601487.0002) is associated with lower body mass index and improved insulin sensitivity among middle-aged and elderly Finns. A significant odds ratio (4.35, P = 0.028) for the association of the pro/pro genotype with type 2 diabetes was observed among Japanese Americans. The PPARG2 ala allele showed decreased binding affinity to the cognate promoter element and reduced ability to transactivate responsive promoters. These findings suggested that the PPARG2 pro12-to-ala polymorphism may contribute to the observed variability in BMI and insulin sensitivity in the general population.

Valve et al. (1999) investigated the frequencies of the pro12-to-ala polymorphism in exon B and the silent CAC478-to-CAT polymorphism in exon 6 of the PPARG gene and their effects on body weight, body composition, and energy expenditure in obese Finnish patients. The frequencies of the ala12 allele in exon B and the CAT478 allele in exon 6 were not significantly different between the obese and population-based control subjects (0.14 vs 0.13 and 0.19 vs 0.21, respectively). The polymorphisms were associated with increased BMI, and the 5 women with both ala12ala and CAT478CAT genotypes were significantly more obese compared with the women having both pro12pro and CAC478CAC genotypes, and they had increased fat mass. The authors concluded that the pro12-to-ala and CAC478-to-CAT polymorphisms in the PPARG gene are associated with severe overweight and increased fat mass among obese women.

Sarraf et al. (1999) identified 4 somatic mutations (1 nonsense, 1 frameshift, and 2 missense) in the PPARG gene among 55 sporadic colon cancers (114500). Each mutation greatly impaired the function of the PPARG protein. The 472delA mutation (601487.0003) resulted in the deletion of the entire ligand binding domain. Q286P (601487.0004) and K319X (601487.0005) retained a total or partial ligand binding domain but lost the ability to activate transcription through a failure to bind to ligands. R288H (601487.0006) showed a normal response to synthetic ligands but greatly decreased transcription and binding when exposed to natural ligands. These data indicated that colon cancer in humans is associated with loss-of-function mutations in the PPARG gene.

Barroso et al. (1999) reported 2 different heterozygous mutations in the ligand-binding domain of PPARG in 3 subjects with severe insulin resistance (604367). In the PPAR-gamma crystal structure, the mutations destabilized helix 12, which mediates transactivation. Consistent with this, both receptor mutants were markedly transcriptionally impaired and, moreover, were

able to inhibit the action of coexpressed wildtype PPAR-gamma in a dominant-negative manner. In addition to insulin resistance, all 3 subjects developed type 2 diabetes mellitus and hypertension at an unusually early age. Barroso et al. (1999) concluded that their findings represented the first germline loss-of-function mutations in PPAR-gamma and provided compelling genetic evidence that this receptor is important in the control of insulin sensitivity, glucose homeostasis, and blood pressure in man.

Kroll et al. (2000) reported that t(2;3)(q13;p25), a translocation identified in a subset of human thyroid follicular carcinomas, results in fusion of the DNA-binding domains of the thyroid transcription factor PAX8 (167415) to domains A to F of PPARG1. PAX8/PPARG1 mRNA and protein were detected in 5 of 8 thyroid follicular carcinomas but not in 20 follicular adenomas, 10 papillary carcinomas, or 10 multinodular hyperplasias. PAX8/PPARG1 inhibited thiazolidinedione-induced transactivation by PPARG1 in a dominant-negative manner. The experiments demonstrated an oncogenic role for PPARG and suggested that PAX8/PPARG1 may be useful in the diagnosis and treatment of thyroid carcinoma.

ANIMAL MODEL

The nuclear hormone receptor PPARG promotes adipogenesis and macrophage differentiation and is a primary pharmacologic target in the treatment of type II diabetes. Barak et al. (1999) showed that PPARG gene knockout in mice resulted in 2 independent lethal phases. Initially, PPARG deficiency interfered with terminal differentiation of the trophoblast and placental vascularization, leading to severe myocardial thinning and death by E10.0. Supplementing PPARG null embryos with wildtype placentas via aggregation with tetraploid embryos corrected the cardiac defect, implicating a previously unrecognized dependence of the developing heart on a functional placenta. A tetraploid-rescued mutant surviving to term exhibited another lethal combination of pathologies, including lipodystrophy and multiple hemorrhages. These findings both confirmed and expanded the current known spectrum of physiologic functions regulated by PPARG.

Kubota et al. (1999) generated homozygous PPARG-deficient mouse embryos, which died at 10.5 to 11.5 days postcoitum due to placental dysfunction. Heterozygous PPARG-deficient mice were protected from the development of insulin resistance due to adipocyte hypertrophy under a high-fat diet. These phenotypes were abrogated by PPARG agonist treatment. Heterozygous PPARG-deficient mice showed overexpression and hypersecretion of

leptin despite the smaller size of adipocytes and decreased fat mass, which may explain these phenotypes at least in part. This study revealed an unpredicted role for PPARG in high-fat diet-induced obesity due to adipocyte hypertrophy and insulin resistance, which requires both alleles of PPARG.

Rosen et al. (1999) demonstrated that mice chimeric for wildtype and PPARG null cells showed little or no contribution of null cells to adipose tissue, whereas most other organs examined did not require PPARG for proper development. In vitro, the differentiation of embryonic stem cells into fat was shown to be dependent on PPARG gene dosage. These data provided direct evidence that PPARG is essential for the formation of fat.

The thiazolidinedione (TZD) class of insulin-sensitizing, antidiabetic drugs interacts with PPAR-gamma. Miles et al. (2000) conducted metabolic studies in PPARG gene knockout mice. Because homozygous PPARG-null mice die in development, they studied glucose metabolism in mice heterozygous for the mutation. They identified no statistically significant differences in body weight, basal glucose, insulin, or free fatty acid levels between the wildtype and heterozygous groups. Nor was there a difference in glucose excursion between the groups of mice during oral glucose tolerance tests. However, insulin concentrations of the wildtype group were greater than those of the heterozygous deficient group, and insulin-induced increase in glucose disposal rate was significantly increased in the heterozygous mice. Likewise, the insulin-induced suppression of hepatic glucose production was significantly greater in the heterozygous mice than in wildtype mice. Taken together, these results indicated that--counterintuitively--although pharmacologic activation of PPAR-gamma improves insulin sensitivity, a similar effect is obtained by genetically reducing the expression levels of the receptor.

ALLELIC VARIANTS (selected examples)

.0001 OBESITY, SEVERE [PPARG, PRO115GLN]

In 4 German subjects with severe obesity (601665), Ristow et al. (1998) identified a pro115-to-gln mutation of the PPAR-gamma-2 gene. Significantly, the mutation was in the codon immediately adjacent to a serine-114 phosphorylation site. The pro115-to-gln mutation occurs in exon 6, which is shared by all 3 forms of PPAR-gamma Wang et al. (1999).

.0002 PPARG2 POLYMORPHISM C/G [PPARG, PRO12ALA]

OBESITY, PROTECTION AGAINST DIABETES MELLITUS, TYPE II, SUSCEPTIBILITY TO, INCLUDED Because the product of the PPARG gene is a nuclear receptor that regulates adipocyte differentiation and possibly lipid metabolism and insulin sensitivity, Yen et al. (1997) screened for mutations in the entire coding region of the PPARG gene in 26 diabetic Caucasians with or without obesity (601665). They found a CCG (pro)-to-GCG (ala) missense mutation at codon 12 (P12A). The allele frequency of the mutation varied from 0.12 in Caucasian Americans to 0.10 in Chinese. Beamer et al. (1998) noted that the amino acid position of the P12A mutation is within the domain of PPAR-gamma-2 that enhances ligand-independent activation, that the substitution of alanine for proline is nonconservative, and that this amino acid change might cause a significant alteration in protein structure. To test the hypothesis that individuals with the variant are at increased genetic risk for obesity and/or insulin resistance, they performed association studies in 2 independently recruited cohorts of unrelated, nondiabetic, adult Caucasian subjects. They found that the P12A mutation was associated with higher BMI in the 2 cohorts, suggesting that the mutation may contribute to genetic susceptibility for the multifactorial disorder of obesity.

Deeb et al. (1998) studied a polymorphism of the PPARG gene, a C-to-G variant that created an HgaI restriction site and predicted the substitution of alanine for proline at position 12 in the PPARG2-specific exon B. In a group of Finnish men and women with a PPARG2 ala allele frequency of 0.12, they found that this allele was associated with lower fasting insulin levels ($P = 0.011$) and BMI ($P = 0.027$) and higher insulin sensitivity ($P = 0.047$). This association was independent of sex. The findings were verified by studies in a group of elderly subjects. They also studied the association of the pro12-to-ala substitution in PPARG2 with type 2 diabetes (125853) in a group of second-generation Japanese-American (Nisei) men and women that included individuals with type 2 diabetes, impaired glucose tolerance, and normal controls. The ala allele was less frequent among subjects with type 2 diabetes (0.022) than among normal controls (0.092). The odds ratio for association of pro/pro with diabetes was significant (4.35, $P = 0.028$), whereas the frequency of the ala allele among impaired glucose tolerance subjects was intermediate (0.039). Deeb et al. (1998) suggested that the lower transactivation capacity of the ala variant of PPARG2 underlies the association of this allele with lower BMI and higher insulin sensitivity. The ala isoform may lead to less efficient stimulation of PPARG target genes and predispose to lower levels of adipose tissue mass accumulation, which in turn may be responsible for improved insulin sensitivity.

Altshuler et al. (2000) evaluated 16 published genetic associations to type 2 diabetes and related subphenotypes using a family-based design to control for population stratification, and replication samples to increase power. They confirmed only 1 association, that of the common pro12-to-ala polymorphism in PPAR-gamma with type 2 diabetes. By analyzing over 3,000 individuals, they found a modest (1.25-fold) but significant (P = 0.002) increase in diabetes risk associated with the more common proline allele (approximately 85% frequency). Because the risk allele occurs at such high frequency, its modest effect translates into a large population-attributable risk--influencing as much as 25% of type 2 diabetes in the general population.

.0003 CANCER OF COLON [PPARG, 1-BP DEL, 472A]

In a sporadic colon cancer (114500) tumor, Sarraf et al. (1999) identified a somatic mutation in the PPARG gene, a 1-bp deletion at nucleotide 472, which resulted in a frameshift.

.0004 CANCER OF COLON [PPARG, GLN286PRO]

In a sporadic colon cancer (114500) tumor, Sarraf et al. (1999) identified a somatic mutation in the PPARG gene, an A-to-G transition at nucleotide 857, which resulted in a gln286-to-pro substitution.

.0005 CANCER OF COLON [PPARG, LYS319TER]

In a sporadic colon cancer (114500), Sarraf et al. (1999) identified a somatic mutation in the PPARG gene, an A-to-T transversion at nucleotide 955, which resulted in a lys319-to-ter substitution.

.0006 CANCER OF COLON [PPARG, ARG288HIS]

In a sporadic colon cancer (114500) tumor, Sarraf et al. (1999) identified a somatic mutation in the PPARG gene, a G-to-A transition at nucleotide 863, which resulted in an arg288-to-his substitution.

.0007 DIABETES MELLITUS, INSULIN-RESISTANT, WITH ACANTHOSIS NIGRICANS AND HYPERTENSION [PPARG, PRO467LEU]

In a patient with severe insulin resistance, type 2 diabetes mellitus, and hypertension (604367) who had been diagnosed in her twenties, Barroso et al. (1999) detected a C-to-T

transition in the PPARG gene resulting in a proline-to-leucine mutation at codon 467 (P467L). Her son, aged 30 years, who also had a history of early-onset diabetes and hypertension, was also heterozygous for the P467L mutation. All other family members, including both parents of the proband, none of whom were known to have diabetes or hypertension, were homozygous for wildtype receptor sequence. Nonpaternity was excluded, indicating a de novo appearance of the mutation in the proband.

.0008 DIABETES MELLITUS, INSULIN-RESISTANT, WITH ACANTHOSIS NIGRICANS AND HYPERTENSION [PPARG, VAL290MET]

In a 15-year-old patient with primary amenorrhea, hirsutism, acanthosis nigricans, elevated blood pressure, and markedly elevated fasting and postprandial insulin levels (604367), Barroso et al. (1999) identified a G-to-A transition in the PPARG gene resulting in a valine-to-methionine mutation at codon 290 (V290M). By age 17 the patient had developed type 2 diabetes and had hypertension which required treatment with beta-blockers. Her clinically unaffected mother and sister were both wildtype at this locus; screening of the deceased father was not possible.

.0009 PPARG POLYMORPHISM C-T [PPARG, 161C-T]

Meirhaeghe et al. (1998) reported a 161C-T substitution in exon 6 of the PPARG gene. Since PPAR-gamma is a transcription factor implicated in adipocyte differentiation and in lipid and glucose metabolism, they analyzed the relationships between this genetic polymorphism and various markers of the obesity phenotype in a representative sample of 820 men and women living in northern France. The frequencies of the C and T alleles were 0.860 and 0.140, respectively. In the whole sample, no association of the polymorphism with the markers tested was observed, but a statistically significant interaction (P less than 0.03) existed between this polymorphism and body mass index (BMI) for plasma leptin levels. Obese subjects bearing at least one T allele had higher plasma leptin levels than subjects who did not. This effect existed in both genders, despite the higher plasma leptin levels observed in women. Thus, for a given leptin level, the BMI was relatively lower in obese subjects carrying at least one T allele than in obese CC homozygotes.

Wang et al. (1999) studied this polymorphism in 647 Australian Caucasian patients aged 65 years or less, with or without angiographically documented coronary artery disease. The

frequencies of the CC, CT, and TT genotypes were 69.8%, 27.7%, and 2.5%, respectively, and the T allele frequency 0.163. These frequencies were in Hardy-Weinberg equilibrium and not different between men and women. Wang et al. (1999) found that the T allele carriers (CT and TT genotypes) had significantly reduced coronary artery disease risk compared to the CC homozygotes, with an odds ratio of 0.457. Association with obesity (601665) was not found in these patients. The authors interpreted this to indicate that the PPARG gene may have a significant role in atherogenesis, independent of obesity and of lipid abnormalities, possibly via a direct local vascular wall effect.

Using a subtractive cloning strategy to identify downstream targets of peroxisome proliferator-activated receptor-gamma (PPARG; 601487), and by screening cDNA libraries, Yoon et al. (2000) isolated mouse and human cDNAs encoding PGAR. The 406-amino acid, 60-kD human PGAR protein, which shares 75% amino acid identity with the mouse protein, is a member of the angiopoietin family of secreted proteins and bears highest similarity to angiopoietin-2 (ANGPT2; 601922). Like other members of this family, PGAR contains a predicted coiled-coil quaternary structure, and the authors hypothesized that PGAR may form multimeric or other higher-order structures. PGAR has a secretory signal peptide, 3 potential N-glycosylation sites, and 4 cysteines that may be available for intramolecular disulfide bonding. Northern blot analysis detected a 2-kb PGAR transcript that was highly enriched in white fat and placenta. In situ hybridization analysis revealed expression of mouse Pgar at low levels in most organs and connective tissue at embryonic day 13.5 (E13.5). Between E15.5 and E18.5, strongest expression of Pgar was in brown fat. Northern blot analysis detected elevated levels of Pgar expression in mouse models of obesity and diabetes. Alterations in nutrition and leptin (164160) administration in mice modulated Pgar expression in vivo. Yoon et al. (2000) demonstrated that PPARG ligand-induced transcription of PGAR follows a rapid time course typical of immediate-early genes and occurs in the absence of protein synthesis. Using a culture model system, they observed that induction of the PGAR transcript coincides with hormone-dependent adipocyte differentiation. Yoon et al. (2000) concluded that PGAR is a bona fide target of PPARG and may have a role in regulation of systemic lipid metabolism or glucose homeostasis.

Kersten et al. (2000) identified mouse Pgar, which they called Fiaf (fasting-induced adipose factor), using a subtractive hybridization assay to identify PPARA (170998) target genes. Northern blot analysis detected expression of Fiaf in mouse white and brown adipose tissue, with weak expression in lung, kidney, and liver. Using a combination of wildtype, Ppara

mutant, and Pparg mutant mice, Kersten et al. (2000) demonstrated that mRNA expression is stimulated by PPARA in liver and by PPARG in white adipose tissue. Expression of Fiaf was upregulated in liver and white adipose tissue during fasting. Western blot analysis showed that the abundance of Fiaf in plasma decreased with high fat feeding, an effect directly opposite that observed with leptin.

By radiation hybrid analysis, Yoon et al. (2000) mapped the PGAR gene to 19p13.3.

The DNA and protein sequences for the novel Angiopoietin-like gene are reported here as CuraGen Acc. No. CG57051-04.

Similarities

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 716 of 733 bases (97%) identical to a gb:GENBANK-ID:AF202636|acc:AF202636.1 mRNA from Homo sapiens (Homo sapiens angiopoietin-like protein PP1158 mRNA, complete cds) (Table 23). The full amino acid sequence of the protein of the invention was found to have 181 of 183 amino acid residues (98%) identical to, and 182 of 183 amino acid residues (99%) similar to, the 406 amino acid residue ptnr:SPTREMBL-ACC:Q9HBV4 protein from Homo sapiens (Human) (ANGIOPOIETIN-LIKE PROTEIN PP1158) (Table 24).

A multiple sequence alignment is given in Table 26, with the protein of the invention being shown on the first line in a ClustalW analysis comparing the protein of the invention with related protein sequences. Please note this sequence represents a splice form of Angiopoietin as indicated in positions 184L to 347G and SNPs: Q24R and G25S.

The presence of identifiable domains in the protein disclosed herein was determined by searches versus domain databases such as Pfam, PROSITE, ProDom, Blocks or Prints and then identified by the Interpro domain accession number. Significant domains are summarized below:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
fibrinogen_C	1/1	184	236 ..	204	272 .]	31.7	4.1e-08

IPR002181; Fibrinogen_C

Fibrinogen [1], the principal protein of vertebrate blood clotting is an hexamer containing two sets of three different chains (alpha, beta, and gamma), linked to each other by disulfide bonds. The N-terminal sections of these three chains are evolutionary related and contain the cysteines that participate in the cross-linking of the chains. However, there is no similarity between the C-terminal part of the alpha chain and that of the beta and gamma chains. The C-terminal part of the beta and gamma chains forms a domain of about 270 amino-acid residues. As shown in the schematic representation this domain contains four conserved cysteines involved in two disulfide bonds. (SEQ ID NO:126)

```
xxxxCxxxxxxxxxxxxxCxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxCxxxxxCxxxxxxxxxxxx
    |             |                                   |     |
    +-------------+                                   +-----+
```

'C': conserved cysteine involved in a disulfide bond.

Such a domain has been recently found in other proteins which are listed below.

Two sea cucumber fibrinogen-like proteins (FReP-A and FReP-B). These are proteins, of about 260 amino acids, which have a fibrinogen beta/gamma C-terminal domain.

In the C-terminus of Drosophila protein scabrous (gene sca). Scabrous is involved in the regulation of neurogenesis in Drosophila and may encode a lateral inhibitor of R8 cells differentiation. In the C-terminus of a mammalian T-cell specific protein of unknown function.

In the C-terminus of a human protein of unknown function which is encoded on the opposite strand of the steroid 21-hydroxylase/complement component C4 gene locus.

The function of this domain is not yet known, but it has been suggested that it could be involved in protein-protein interactions.

This indicates that the sequence of the invention has properties similar to those of other proteins known to contain this/these domain(s) and similar to the properties of these domains.

Chromosomal information:

The Angiopoietin-like gene disclosed in this invention maps to chromosome 19p13.3. This assignment was made using mapping information associated with genomic clones, public genes and ESTs sharing sequence identity with the disclosed sequence and CuraGen Corporation's Electronic Northern bioinformatic tool.

Tissue expression

The Angiopoietin-like gene disclosed in this invention is expressed in at least the following tissues: Adipose, Heart, Aorta, Coronary Artery, Umbilical Vein, Adrenal Gland/Suprarenal gland, Pancreas, Islets of Langerhans, Thyroid, Pineal Gland, Parotid Salivary glands, Liver, Small Intestine, Duodenum, Colon, Bone Marrow, Lymph node, Bone, Cartilage, Synovium/Synovial membrane, Skeletal Muscle, Brain, Thalamus, Pituitary Gland, Amygdala, Hippocampus, Spinal Chord, Mammary gland/Breast, Ovary, Placenta, Uterus, Vulva, Prostate, Testis, Lung, Kidney, Retina, Skin, Foreskin. Expression information was derived from the tissue sources of the sequences that were included in the derivation of the sequence of CuraGen Acc. No. CG57051-04.

Cellular Localization and Sorting

The PSORT, SignalP and hydropathy profile for the Angiopoietin-like protein are shown in Table 27. Although PSORT suggests that the Angiopoietin-like protein may be localized in the cytoplasm, the protein of CuraGen Acc. No. CG57051-04 predicted here is similar to the Fibrinogen family, some members of which are secreted. Therefore it is likely that this novel Angiopoietin-like protein is localized to the same sub-cellular compartment.

Functional Variants and Homologs

The novel nucleic acid of the invention encoding a Angiopoietin-like protein includes the nucleic acid whose sequence is provided in Figure 20, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Fig. 1 while still encoding a protein that maintains its Angiopoietin-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to the sequence of CuraGen Acc. No. CG57051-04, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in

therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 3% of the bases may be so changed.

The novel protein of the invention includes the Angiopoietin-like protein whose sequence is provided in Figure 20. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Figure 20 while still encoding a protein that maintains its Angiopoietin-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 2% of the amino acid residues may be so changed.

Chimeric and Fusion Proteins

The present invention includes chimeric or fusion proteins of the Angiopoietin-like protein, in which the Angiopoietin-like protein of the present invention is joined to a second polypeptide or protein that is not substantially homologous to the present novel protein. The second polypeptide can be fused to either the amino-terminus or carboxyl-terminus of the present CG57051-04 polypeptide. In certain embodiments a third nonhomologous polypeptide or protein may also be fused to the novel Angiopoietin-like protein such that the second nonhomologous polypeptide or protein is joined at the amino terminus, and the third nonhomologous polypeptide or protein is joined at the carboxyl terminus, of the CG57051-04 polypeptide. Examples of nonhomologous sequences that may be incorporated as either a second or third polypeptide or protein include glutathione S-transferase, a heterologous signal sequence fused at the amino terminus of the Angiopoietin-like protein, an immunoglobulin sequence or domain, a serum protein or domain thereof (such as a serum albumin), an antigenic epitope, and a specificity motif such as $(His)_6$.

The invention further includes nucleic acids encoding any of the chimeric or fusion proteins described in the preceding paragraph.

Antibodies

The invention further encompasses antibodies and antibody fragments, such as Fab, $(Fab)_2$ or single chain FV constructs, that bind immunospecifically to any of the proteins of the invention. Also encompassed within the invention are peptides and polypeptides comprising sequences having high binding affinity for any of the proteins of the invention, including such

peptides and polypeptides that are fused to any carrier particle (or biologically expressed on the surface of a carrier) such as a bacteriophage particle.

Uses of the Compositions of the Invention

The protein similarity information, expression pattern, cellular localization, and map location for the protein and nucleic acid disclosed herein suggest that this Angiopoietin-like protein may have important structural and/or physiological functions characteristic of the Fibrinogen family. Therefore, the nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications and as a research tool. These include serving as a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed. These also include potential therapeutic applications such as the following: (i) a protein therapeutic, (ii) a small molecule drug target, (iii) an antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) a nucleic acid useful in gene therapy (gene delivery/gene ablation), (v) an agent promoting tissue regeneration *in vitro* and *in vivo*, and (vi) a biological defense weapon.

The nucleic acids and proteins of the invention have applications in the diagnosis and/or treatment of various diseases and disorders. For example, the compositions of the present invention will have efficacy for the treatment of patients suffering from: type II diabetes, obesity, colon cancer, diabetes mellitus, insulin-resistant, with acanthosis nigricans and hypertension, 3-methylglutaconicaciduria, type III; Cone-rod retinal dystrophy-2;DNA ligase I deficiency; Glutaricaciduria, type IIB Liposarcoma; Myotonic dystrophy as well as other diseases, disorders and conditions.

These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in diagnostic and/or therapeutic methods.

Table 23. BLASTN search using CuraGen Acc. No. CG57051-04.

>gb:GENBANK-ID:AF202636|acc:AF202636.1 Homo sapiens angiopoietin-like protein PP1158 mRNA, complete cds - Homo sapiens, 1943 bp.
Length = 1943 (SEQ ID NO:79)

Plus Strand HSPs:

Score = 3468 (520.3 bits), Expect = 7.8e-202, Sum P(2) = 7.8e-202
Identities = 716/733 (97%), Positives = 716/733 (97%), Strand = Plus / Plus

```
Query:    2 GCGGATCCTCACACGACTGTGATCCGATTCTTTCCAGCGGCTTCTGCAACCAAGCGGGTC 61
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   20 GCGGATCCTCACACGACTGTGATCCGATTCTTTCCAGCGGCTTCTGCAACCAAGCGGGTC 79

Query:   62 TTACCCCCGGTCCTCCGCGTCTCCAGTCCTCGCACCTGGAACCCCAACGTCCCCGAGAGT 121
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   80 TTACCCCCGGTCCTCCGCGTCTCCAGTCCTCGCACCTGGAACCCCAACGTCCCCGAGAGT 139

Query:  122 CCCCGAATCCCCGCTCCCAGGCTACCTAAGAGGATGAGCGGTGCTCCGACGGCCGGGGCA 181
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  140 CCCCGAATCCCCGCTCCCAGGCTACCTAAGAGGATGAGCGGTGCTCCGACGGCCGGGGCA 199

Query:  182 GCCCTGATGCTCTGCGCCGCCACCGCCGTGCTACTGAGCGCT-AGATCTGGACCCGTGCA 240
            ||||||||||||||||||||||||||||||||||||||||||| ||  |  |||||||||||
Sbjct:  200 GCCCTGATGCTCTGCGCCGCCACCGCCGTGCTACTGAGCGCTCAGGGC-GGACCCGTGCA 258

Query:  241 GTCCAAGTCGCCGCGCTTTGCGTCCTGGGACGAGATGAATGTCCTGGCGCACGGACTCCT 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  259 GTCCAAGTCGCCGCGCTTTGCGTCCTGGGACGAGATGAATGTCCTGGCGCACGGACTCCT 318

Query:  301 GCAGCTCGGCCAGGGGCTGCGCGAACACGCGGAGCGCACCCGCAGTCAGCTGAGCGCGCT 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  319 GCAGCTCGGCCAGGGGCTGCGCGAACACGCGGAGCGCACCCGCAGTCAGCTGAGCGCGCT 378

Query:  361 GGAGCGGCGCCTGAGCGCGTGCGGGTCCGCCTGTCAGGGAACCGAGGGGTCCACCGACCT 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  379 GGAGCGGCGCCTGAGCGCGTGCGGGTCCGCCTGTCAGGGAACCGAGGGGTCCACCGACCT 438

Query:  421 CCCGTTAGCCCCTGAGAGCCGGGTGGACCCTGAGGTCCTTCACAGCCTGCAGACACAACT 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  439 CCCGTTAGCCCCTGAGAGCCGGGTGGACCCTGAGGTCCTTCACAGCCTGCAGACACAACT 498

Query:  481 CAAGGCTCAGAACAGCAGGATCCAGCAACTCTTCCACAAGGTGGCCCAGCAGCAGCGGCA 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  499 CAAGGCTCAGAACAGCAGGATCCAGCAACTCTTCCACAAGGTGGCCCAGCAGCAGCGGCA 558

Query:  541 CCTGGAGAAGCAGCACCTGCGAATTCAGCATCTGCAAAGCCAGTTTGGCCTCCTGGACCA 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  559 CCTGGAGAAGCAGCACCTGCGAATTCAGCATCTGCAAAGCCAGTTTGGCCTCCTGGACCA 618

Query:  601 CAAGCACCTAGACCATGAGGTGGCCAAGCCTGCCCGAAGAAAGAGGCTGCCCGAGATGGC 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  619 CAAGCACCTAGACCATGAGGTGGCCAAGCCTGCCCGAAGAAAGAGGCTGCCCGAGATGGC 678

Query:  661 CCAGCCAGTTGACCCGGCTCACAATGTCAGCCGCCTGCACCGAG-GCTGGTGGTTTGGCA 719
            ||||||||||||||||||||||||||||||||||||||||||   ||    || ||| ||
Sbjct:  679 CCAGCCAGTTGACCCGGCTCACAATGTCAGCCGCCTGCACCGGCTGCCCAGGGATTGCCA 738

Query:  720 CCTGCAGCCATTCCA 734
               | |||  |||||
Sbjct:  739 G--G-AGCTGTTCCA 750
```

Score = 1182 (177.3 bits), Expect = 7.8e-202, Sum P(2) = 7.8e-202
Identities = 242/245 (98%), Positives = 242/245 (98%), Strand = Plus / Plus

```
Query:  693 GCCTGCACCG-AGGCTGGTGGTTTGGCACCTGCAGCCATTCCAACCTCAACGGCCAGTAC 751
            |||| | | | |||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1203 GCCT-CTCTGGAGGCTGGTGGTTTGGCACCTGCAGCCATTCCAACCTCAACGGCCAGTAC 1261

Query:  752 TTCCGCTCCATCCCACAGCAGCGGCAGAAGCTTAAGAAGGGAATCTTCTGGAAGACCTGG 811
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1262 TTCCGCTCCATCCCACAGCAGCGGCAGAAGCTTAAGAAGGGAATCTTCTGGAAGACCTGG 1321

Query:  812 CGGGGCCGCTACTACCCGCTGCAGGCCACCACCATGTTGATCCAGCCCATGGCAGCAGAG 871
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1322 CGGGGCCGCTACTACCCGCTGCAGGCCACCACCATGTTGATCCAGCCCATGGCAGCAGAG 1381

Query:  872 GCAGCCTCCTAGCGTCCTGGCTGGGCCTGGTCCCAGGCCCACGAAAGACGGTGACTCTTG 931
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1382 GCAGCCTCCTAGCGTCCTGGCTGGGCCTGGTCCCAGGCCCACGAAAGACGGTGACTCTTG 1441
```

```
Query:   932 GCTCTG 937
             ||||||
Sbjct:  1442 GCTCTG 1447
```

Table 24. BLASTP search using the protein of CuraGen Acc. No. CG57051-04.

>ptnr:SPTREMBL-ACC:Q9HBV4 ANGIOPOIETIN-LIKE PROTEIN PP1158 - Homo sapiens (Human), 406 aa. (SEQ ID NO:80)
Length = 406

Score = 929 (327.0 bits), Expect = 4.4e-126, Sum P(2) = 4.4e-126
Identities = 181/183 (98%), Positives = 182/183 (99%)

```
Query:     1 MSGAPTAGAALMLCAATAVLLSARSGPVQSKSPRFASWDEMNVLAHGLLQLGQGLREHAE 60
             |||||||||||||||||||||||+ |||||||||||||||||||||||||||||||||||
Sbjct:     1 MSGAPTAGAALMLCAATAVLLSAQGGPVQSKSPRFASWDEMNVLAHGLLQLGQGLREHAE 60

Query:    61 RTRSQLSALERRLSACGSACQGTEGSTDLPLAPESRVDPEVLHSLQTQLKAQNSRIQQLF 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    61 RTRSQLSALERRLSACGSACQGTEGSTDLPLAPESRVDPEVLHSLQTQLKAQNSRIQQLF 120

Query:   121 HKVAQQQRHLEKQHLRIQHLQSQFGLLDHKHLDHEVAKPARRKRLPEMAQPVDPAHNVSR 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   121 HKVAQQQRHLEKQHLRIQHLQSQFGLLDHKHLDHEVAKPARRKRLPEMAQPVDPAHNVSR 180

Query:   181 LHR 183
             |||
Sbjct:   181 LHR 183
```

```
 Score = 333 (117.2 bits), Expect = 4.4e-126, Sum P(2) = 4.4e-126
 Identities = 60/62 (96%), Positives = 60/62 (96%)
```

```
Query:   181 LHRGWWFGTCSHSNLNGQYFRSIPQQRQKLKKGIFWKTWRGRYYPLQATTMLIQPMAAEA 240
             |  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   345 LSGGWWFGTCSHSNLNGQYFRSIPQQRQKLKKGIFWKTWRGRYYPLQATTMLIQPMAAEA 404

Query:   241 AS 242
             ||
Sbjct:   405 AS 406
```

Score = 49 (17.2 bits), Expect = 2.4e-33, Sum P(2) = 2.4e-33
Identities = 14/40 (35%), Positives = 20/40 (50%)

```
Query:     1 MSGAPTAGAALMLCAATAVLLSARSGPVQSKSPRFASWDE 40
             + |  ||  +| | |  |  | | + |    |  |++||+
Sbjct:   293 LGGEDTA-YSLQLTAPVAGQLGATTVPPSGLSVPFSTWDQ 331
```

Table 25. BLASTN identity search of CuraGen Corporation's Human SeqCalling database using CuraGen Acc. No. CG57051-04.

>s3aq:230527544 , 2394 bp. (SEQ ID NO:81)
Length = 2394

Minus Strand HSPs:

Score = 3468 (520.3 bits), Expect = 1.2e-202, Sum P(2) = 1.2e-202
Identities = 716/733 (97%), Positives = 716/733 (97%), Strand = Minus / Plus

```
Query:   734 TGGAATGGCTGCAGGTGCCAAACCACCAGCCTC-GGTGCAGGCGGCTGACATTGTGAGCC 676
             |||||  ||| | ||   ||| ||    ||  | ||||||||||||||||||||||||||
Sbjct:  1645 TGGAACAGCTCCTGG---CAATCCCTGGGCAGCCGGTGCAGGCGGCTGACATTGTGAGCC 1701
```

```
Query:   675 GGGTCAACTGGCTGGGCCATCTCGGGCAGCCTCTTTCTTCGGGCAGGCTTGGCCACCTCA 616
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1702 GGGTCAACTGGCTGGGCCATCTCGGGCAGCCTCTTTCTTCGGGCAGGCTTGGCCACCTCA 1761

Query:   615 TGGTCTAGGTGCTTGTGGTCCAGGAGGCCAAACTGGCTTTGCAGATGCTGAATTCGCAGG 556
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1762 TGGTCTAGGTGCTTGTGGTCCAGGAGGCCAAACTGGCTTTGCAGATGCTGAATTCGCAGG 1821

Query:   555 TGCTGCTTCTCCAGGTGCCGCTGCTGCTGGGCCACCTTGTGGAAGAGTTGCTGGATCCTG 496
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1822 TGCTGCTTCTCCAGGTGCCGCTGCTGCTGGGCCACCTTGTGGAAGAGTTGCTGGATCCTG 1881

Query:   495 CTGTTCTGAGCCTTGAGTTGTGTCTGCAGGCTGTGAAGGACCTCAGGGTCCACCCGGCTC 436
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1882 CTGTTCTGAGCCTTGAGTTGTGTCTGCAGGCTGTGAAGGACCTCAGGGTCCACCCGGCTC 1941

Query:   435 TCAGGGGCTAACGGGAGGTCGGTGGACCCCTCGGTTCCCTGACAGGCGGACCCGCACGCG 376
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1942 TCAGGGGCTAACGGGAGGTCGGTGGACCCCTCGGTTCCCTGACAGGCGGACCCGCACGCG 2001

Query:   375 CTCAGGCGCCGCTCCAGCGCGCTCAGCTGACTGCGGGTGCGCTCCGCGTGTTCGCGCAGC 316
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  2002 CTCAGGCGCCGCTCCAGCGCGCTCAGCTGACTGCGGGTGCGCTCCGCGTGTTCGCGCAGC 2061

Query:   315 CCCTGGCCGAGCTGCAGGAGTCCGTGCGCCAGGACATTCATCTCGTCCCAGGACGCAAAG 256
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  2062 CCCTGGCCGAGCTGCAGGAGTCCGTGCGCCAGGACATTCATCTCGTCCCAGGACGCAAAG 2121

Query:   255 CGCGGCGACTTGGACTGCACGGGTCCAGATCT-AGCGCTCAGTAGCACGGCGGTGGCGGC 197
             |||||||||||||||||||||||||| |  || |||||||||||||||||||||||||||
Sbjct:  2122 CGCGGCGACTTGGACTGCACGGGTCC-GCCCTGAGCGCTCAGTAGCACGGCGGTGGCGGC 2180

Query:   196 GCAGAGCATCAGGGCTGCCCCGGCCGTCGGAGCACCGCTCATCCTCTTAGGTAGCCTGGG 137
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  2181 GCAGAGCATCAGGGCTGCCCCGGCCGTCGGAGCACCGCTCATCCTCTTAGGTAGCCTGGG 2240

Query:   136 AGCGGGGATTCGGGGACTCTCGGGGACGTTGGGGTTCCAGGTGCGAGGACTGGAGACGCG 77
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  2241 AGCGGGGATTCGGGGACTCTCGGGGACGTTGGGGTTCCAGGTGCGAGGACTGGAGACGCG 2300

Query:    76 GAGGACCGGGGGTAAGACCCGCTTGGTTGCAGAAGCCGCTGGAAAGAATCGGATCACAGT 17
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  2301 GAGGACCGGGGGTAAGACCCGCTTGGTTGCAGAAGCCGCTGGAAAGAATCGGATCACAGT 2360

Query:    16 CGTGTGAGGATCCGC 2
             |||||||||||||||
Sbjct:  2361 CGTGTGAGGATCCGC 2375
```

Score = 1182 (177.3 bits), Expect = 1.2e-202, Sum P(2) = 1.2e-202 **(SEQ ID NO:127)**
Identities = 242/245 (98%), Positives = 242/245 (98%), Strand = Minus / Plus

```
Query:   937 CAGAGCCAAGAGTCACCGTCTTTCGTGGGCCTGGGACCAGGCCCAGCCAGGACGCTAGGA 878
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   948 CAGAGCCAAGAGTCACCGTCTTTCGTGGGCCTGGGACCAGGCCCAGCCAGGACGCTAGGA 1007

Query:   877 GGCTGCCTCTGCTGCCATGGGCTGGATCAACATGGTGGTGGCCTGCAGCGGGTAGTAGCG 818
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1008 GGCTGCCTCTGCTGCCATGGGCTGGATCAACATGGTGGTGGCCTGCAGCGGGTAGTAGCG 1067

Query:   817 GCCCCGCCAGGTCTTCCAGAAGATTCCCTTCTTAAGCTTCTGCCGCTGCTGTGGGATGGA 758
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1068 GCCCCGCCAGGTCTTCCAGAAGATTCCCTTCTTAAGCTTCTGCCGCTGCTGTGGGATGGA 1127

Query:   757 GCGGAAGTACTGGCCGTTGAGGTTGGAATGGCTGCAGGTGCCAAACCACCAGCCTCG-GT 699
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  | 
Sbjct:  1128 GCGGAAGTACTGGCCGTTGAGGTTGGAATGGCTGCAGGTGCCAAACCACCAGCCTCCAGA 1187

Query:   698 GCAGGC 693
             | ||||
Sbjct:  1188 G-AGGC 1192
```

>s3aq:218296061 , 1862 bp. (SEQ ID NO:82)
Length = 1862

Minus Strand HSPs:

Score = 3444 (516.7 bits), Expect = 1.8e-201, Sum P(2) = 1.8e-201
Identities = 714/733 (97%), Positives = 714/733 (97%), Strand = Minus / Plus

```
Query:   734 TGGAATGGCTGCAGGTGCCAAACCACCAGCCTC-GGTGCAGGCGGCTGACATTGTGAGCC 676
             |||||  ||| | ||   ||| ||    ||  | ||||||||||||||||||||||||||
Sbjct:  1133 TGGAACAGCTCCTGG---CAATCCCTGGGCAGCCGGTGCAGGCGGCTGACATTGTGAGCC 1189

Query:   675 GGGTCAACTGGCTGGGCCATCTCGGGCAGCCTCTTTCTTCGGGCAGGCTTGGCCACCTCA 616
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1190 GGGTCAACTGGCTGGGCCATCTCGGGCAGCCTCTTTCTTCGGGCAGGCTTGGCCACCTCA 1249

Query:   615 TGGTCTAGGTGCTTGTGGTCCAGGAGGCCAAACTGGCTTTGCAGATGCTGAATTCGCAGG 556
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1250 TGGTCTAGGTGCTTGTGGTCCAGGAGGCCAAACTGGCTTTGCAGATGCTGAATTCGCAGG 1309

Query:   555 TGCTGCTTCTCCAGGTGCCGCTGCTGCTGGGCCACCTTGTGGAAGAGTTGCTGGATCCTG 496
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1310 TGCTGCTTCTCCAGGTGCCGCTGCTGCTGGGCCACCTTGTGGAAGAGTTGCTGGATCCTG 1369

Query:   495 CTGTTCTGAGCCTTGAGTTGTGTCTGCAGGCTGTGAAGGACCTCAGGGTCCACCCGGCTC 436
             |||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Sbjct:  1370 CTGTTCTGAGCCTTGAGTTGTGTCTGCAGGCTGTGAAGGACCCCAGGGTCCACCCGGCTC 1429

Query:   435 TCAGGGGCTAACGGGAGGTCGGTGGACCCCTCGGTTCCCTGACAGGCGGACCCGCACGCG 376
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1430 TCAGGGGCTAACGGGAGGTCGGTGGACCCCTCGGTTCCCTGACAGGCGGACCCGCACGCG 1489

Query:   375 CTCAGGCGCCGCTCCAGCGCGCTCAGCTGACTGCGGGTGCGCTCCGCGTGTTCGCGCAGC 316
             ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1490 CTCAGGCGC-GCTCCAGCGCGCTCAGCTGACTGCGGGTGCGCTCCGCGTGTTCGCGCAGC 1548

Query:   315 CCCTGGCCGAGCTGCAGGAGTCCGTGCGCCAGGACATTCATCTCGTCCCAGGACGCAAAG 256
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1549 CCCTGGCCGAGCTGCAGGAGTCCGTGCGCCAGGACATTCATCTCGTCCCAGGACGCAAAG 1608

Query:   255 CGCGGCGACTTGGACTGCACGGGTCCAGATCT-AGCGCTCAGTAGCACGGCGGTGGCGGC 197
             |||||||||||||||||||||||||| |  || |||||||||||||||||||||||||||
Sbjct:  1609 CGCGGCGACTTGGACTGCACGGGTCC-GCCCTGAGCGCTCAGTAGCACGGCGGTGGCGGC 1667

Query:   196 GCAGAGCATCAGGGCTGCCCCGGCCGTCGGAGCACCGCTCATCCTCTTAGGTAGCCTGGG 137
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1668 GCAGAGCATCAGGGCTGCCCCGGCCGTCGGAGCACCGCTCATCCTCTTAGGTAGCCTGGG 1727

Query:   136 AGCGGGGATTCGGGGACTCTCGGGGACGTTGGGGTTCCAGGTGCGAGGACTGGAGACGCG 77
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1728 AGCGGGGATTCGGGGACTCTCGGGGACGTTGGGGTTCCAGGTGCGAGGACTGGAGACGCG 1787

Query:    76 GAGGACCGGGGGTAAGACCCGCTTGGTTGCAGAAGCCGCTGGAAAGAATCGGATCACAGT 17
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1788 GAGGACCGGGGGTAAGACCCGCTTGGTTGCAGAAGCCGCTGGAAAGAATCGGATCACAGT 1847

Query:    16 CGTGTGAGGATCCGC 2
             |||||||||||||||
Sbjct:  1848 CGTGTGAGGATCCGC 1862
```

Score = 1182 (177.3 bits), Expect = 1.8e-201, Sum P(2) = 1.8e-201 **(SEQ ID NO:128)**
Identities = 242/245 (98%), Positives = 242/245 (98%), Strand = Minus / Plus

```
Query:   937 CAGAGCCAAGAGTCACCGTCTTTCGTGGGCCTGGGACCAGGCCCAGCCAGGACGCTAGGA 878
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   436 CAGAGCCAAGAGTCACCGTCTTTCGTGGGCCTGGGACCAGGCCCAGCCAGGACGCTAGGA 495

Query:   877 GGCTGCCTCTGCTGCCATGGGCTGGATCAACATGGTGGTGGCCTGCAGCGGGTAGTAGCG 818
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   496 GGCTGCCTCTGCTGCCATGGGCTGGATCAACATGGTGGTGGCCTGCAGCGGGTAGTAGCG 555

Query:   817 GCCCCGCCAGGTCTTCCAGAAGATTCCCTTCTTAAGCTTCTGCCGCTGCTGTGGGATGGA 758
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   556 GCCCCGCCAGGTCTTCCAGAAGATTCCCTTCTTAAGCTTCTGCCGCTGCTGTGGGATGGA 615

Query:   757 GCGGAAGTACTGGCCGTTGAGGTTGGAATGGCTGCAGGTGCCAAACCACCAGCCTCG-GT 699
```

```
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||  |
Sbjct:  616 GCGGAAGTACTGGCCGTTGAGGTTGGAATGGCTGCAGGTGCCAAACCACCAGCCTCCAGA 675

Query:  698 GCAGGC 693
            | ||||
Sbjct:  676 G-AGGC 680
```

>s3aq:217940431 Category E: , 530 bp. (SEQ ID NO:83)
Length = 530

Minus Strand HSPs:

Score = 1800 (270.1 bits), Expect = 1.2e-75, P = 1.2e-75
Identities = 384/403 (95%), Positives = 384/403 (95%), Strand = Minus / Plus

```
Query:  631 AGGCTTGGCCACC-TCATGGTCTAGGTG-CTT-GTGGTCCAG-GAGGCCAAACTGGCTTT 576
            || | ||| | || ||| | || | ||| ||  ||   |||  |||||||  ||||||||
Sbjct:  128 AGCCCTGGTCCCCGTCA-G-TCAATGTGACTGAGTCCGCCATTGAGGCCAGTCTGGCTTT 185

Query:  575 GCAGATGCTGAATTCGCAGGTGCTGCTTCTCCAGGTGCCGCTGCTGCTGGGCCACCTTGT 516
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  186 GCAGATGCTGAATTCGCAGGTGCTGCTTCTCCAGGTGCCGCTGCTGCTGGGCCACCTTGT 245

Query:  515 GGAAGAGTTGCTGGATCCTGCTGTTCTGAGCCTTGAGTTGTGTCTGCAGGCTGTGAAGGA 456
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  246 GGAAGAGTTGCTGGATCCTGCTGTTCTGAGCCTTGAGTTGTGTCTGCAGGCTGTGAAGGA 305

Query:  455 CCTCAGGGTCCACCCGGCTCTCAGGGGCTAACGGGAGGTCGGTGGACCCCTCGGTTCCCT 396
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  306 CCTCAGGGTCCACCCGGCTCTCAGGGGCTAACGGGAGGTCGGTGGACCCCTCGGTTCCCT 365

Query:  395 GACAGGCGGACCCGCACGCGCTCAGGCGCCGCTCCAGCGCGCTCAGCTGACTGCGGGTGC 336
            ||||||||||||||||||||||||||||||| | ||||||||||||||||||||||||||
Sbjct:  366 GACAGGCGGACCCGCACGCGCTCAGGCGCCGTTTCAGCGCGCTCAGCTGACTGCGGGTGC 425

Query:  335 GCTCCGCGTGTTCGCGCAGCCCCTGGCCGAGCTGCAGGAGTCCGTGCGCCAGGACATTCA 276
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  426 GCTCCGCGTGTTCGCGCAGCCCCTGGCCGAGCTGCAGGAGTCCGTGCGCCAGGACATTCA 485

Query:  275 TCTCGTCCCAGGACGCAAAGCGCGGCGACTTGGACTGCACGGGTC 231
            |||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  486 TCTCGTCCCAGGACGCAAAGCGCGGCGACTTGGACTGCACGGGTC 530
```

>s3aq:230121563 , 788 bp. (SEQ ID NO:84)
Length = 788

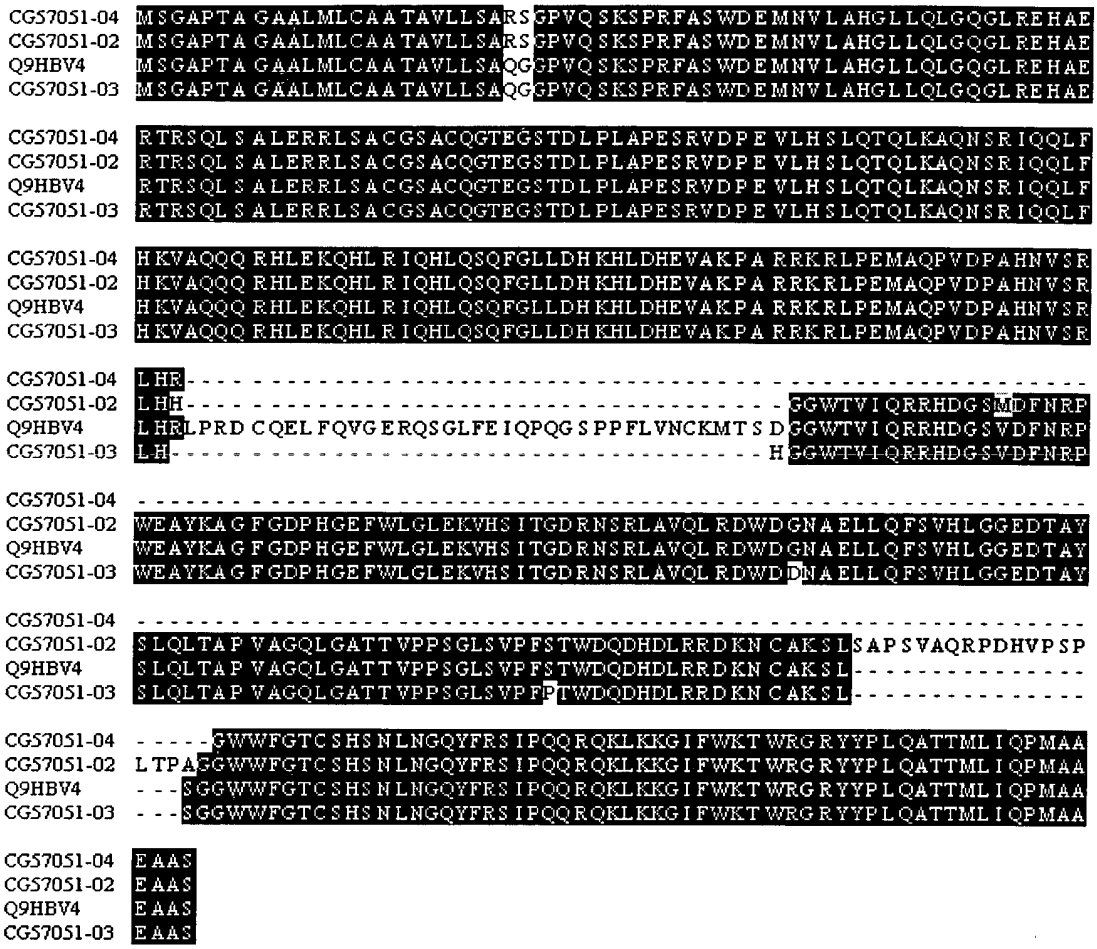
Minus Strand HSPs:

Score = 1182 (177.3 bits), Expect = 6.4e-48, P = 6.4e-48
Identities = 242/245 (98%), Positives = 242/245 (98%), Strand = Minus / Plus

```
Query:  937 CAGAGCCAAGAGTCACCGTCTTTCGTGGGCCTGGGACCAGGCCCAGCCAGGACGCTAGGA 878
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  171 CAGAGCCAAGAGTCACCGTCTTTCGTGGGCCTGGGACCAGGCCCAGCCAGGACGCTAGGA 230

Query:  877 GGCTGCCTCTGCTGCCATGGGCTGGATCAACATGGTGGTGGCCTGCAGCGGGTAGTAGCG 818
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  231 GGCTGCCTCTGCTGCCATGGGCTGGATCAACATGGTGGTGGCCTGCAGCGGGTAGTAGCG 290

Query:  817 GCCCCGCCAGGTCTTCCAGAAGATTCCCTTCTTAAGCTTCTGCCGCTGCTGTGGGATGGA 758
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  291 GCCCCGCCAGGTCTTCCAGAAGATTCCCTTCTTAAGCTTCTGCCGCTGCTGTGGGATGGA 350

Query:  757 GCGGAAGTACTGGCCGTTGAGGTTGGAATGGCTGCAGGTGCCAAACCACCAGCCTCG-GT 699
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  |
Sbjct:  351 GCGGAAGTACTGGCCGTTGAGGTTGGAATGGCTGCAGGTGCCAAACCACCAGCCTCCAGA 410

Query:  698 GCAGGC 693
            | ||||
Sbjct:  411 G-AGGC 415
```

```
>s3aq:217939973 , 631 bp. (SEQ ID NO:85)
       Length = 631

 Minus Strand HSPs:

 Score = 1182 (177.3 bits), Expect = 8.0e-48, P = 8.0e-48
 Identities = 242/245 (98%), Positives = 242/245 (98%), Strand = Minus / Plus

Query:   937 CAGAGCCAAGAGTCACCGTCTTTCGTGGGCCTGGGACCAGGCCCAGCCAGGACGCTAGGA 878
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   105 CAGAGCCAAGAGTCACCGTCTTTCGTGGGCCTGGGACCAGGCCCAGCCAGGACGCTAGGA 164

Query:   877 GGCTGCCTCTGCTGCCATGGGCTGGATCAACATGGTGGTGGCCTGCAGCGGGTAGTAGCG 818
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   165 GGCTGCCTCTGCTGCCATGGGCTGGATCAACATGGTGGTGGCCTGCAGCGGGTAGTAGCG 224

Query:   817 GCCCCGCCAGGTCTTCCAGAAGATTCCCTTCTTAAGCTTCTGCCGCTGCTGTGGGATGGA 758
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   225 GCCCCGCCAGGTCTTCCAGAAGATTCCCTTCTTAAGCTTCTGCCGCTGCTGTGGGATGGA 284

Query:   757 GCGGAAGTACTGGCCGTTGAGGTTGGAATGGCTGCAGGTGCCAAACCACCAGCCTCG-GT 699
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  | 
Sbjct:   285 GCGGAAGTACTGGCCGTTGAGGTTGGAATGGCTGCAGGTGCCAAACCACCAGCCTCCAGA 344

Query:   698 GCAGGC 693
             | ||||
Sbjct:   345 G-AGGC 349

>s3aq:217939964 , 328 bp. (SEQ ID NO:86)
       Length = 328

 Plus Strand HSPs:

 Score = 777 (116.6 bits), Expect = 3.0e-29, P = 3.0e-29
 Identities = 157/159 (98%), Positives = 157/159 (98%), Strand = Plus / Plus

Query:   779 AAGCTTAAGAAGGGAATCTTCTGGAAGACCTGGCGGGGCCGCTACTACCCGCTGCAGGCC 838
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:     1 AAGCTTAAGAAGGGAATCTTCTGGAAGACCTGGCGGGGCCGCTACTACCCGCTGCAGGCC 60

Query:   839 ACCACCATGTTGATCCAGCCCATGGCAGCAGAGGCAGCCTCCTAGCGTCCTGGCTGGGCC 898
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    61 ACCACCATGTTGATCCAGCCCATGGCAGCAGAGGCAGCCTCCTAGCGTCCTGGCTGGGCC 120

Query:   899 TGGTCCCAGGCCCACGAAAGACGGTGACTCTTGGCTCTG 937
             |||||||||||| |||||||||||||||||||||||| |
Sbjct:   121 TGGTCCCAGGCCAACGAAAGACGGTGACTCTTGGCTCCG 159
```

Table 26. ClustalW alignment of CG57051-04 protein with related proteins.



Information for the ClustalW proteins:

Accno	Common Name	Length
CG57051-04 (SEQ ID NO:51)	novel Angiopoietin-like protein	242
CG57051-02 (SEQ ID NO:55)	Angiopoietin Related protein / PPAR-gamma	386
Q9HBV4 (SEQ ID NO:80)	ANGIOPOIETIN-LIKE PROTEIN PP1158.	406
CG57051-03 (SEQ ID NO:57)	Angiopoietin-like protein- isoform 3	368

In the alignment shown above, black outlined amino acid residues indicate residues identically conserved between sequences (i.e., residues that may be required to preserve structural or functional properties); amino acid residues with a gray background are similar to one another between sequences, possessing comparable physical and/or chemical properties

without altering protein structure or function (e.g. the group L,V, I, and M may be considered similar); and amino acid residues with a white background are neither conserved nor similar between sequences.

Table 27. PSORT, SignalP and hydropathy results for CuraGen Acc. No. CG57051-04.

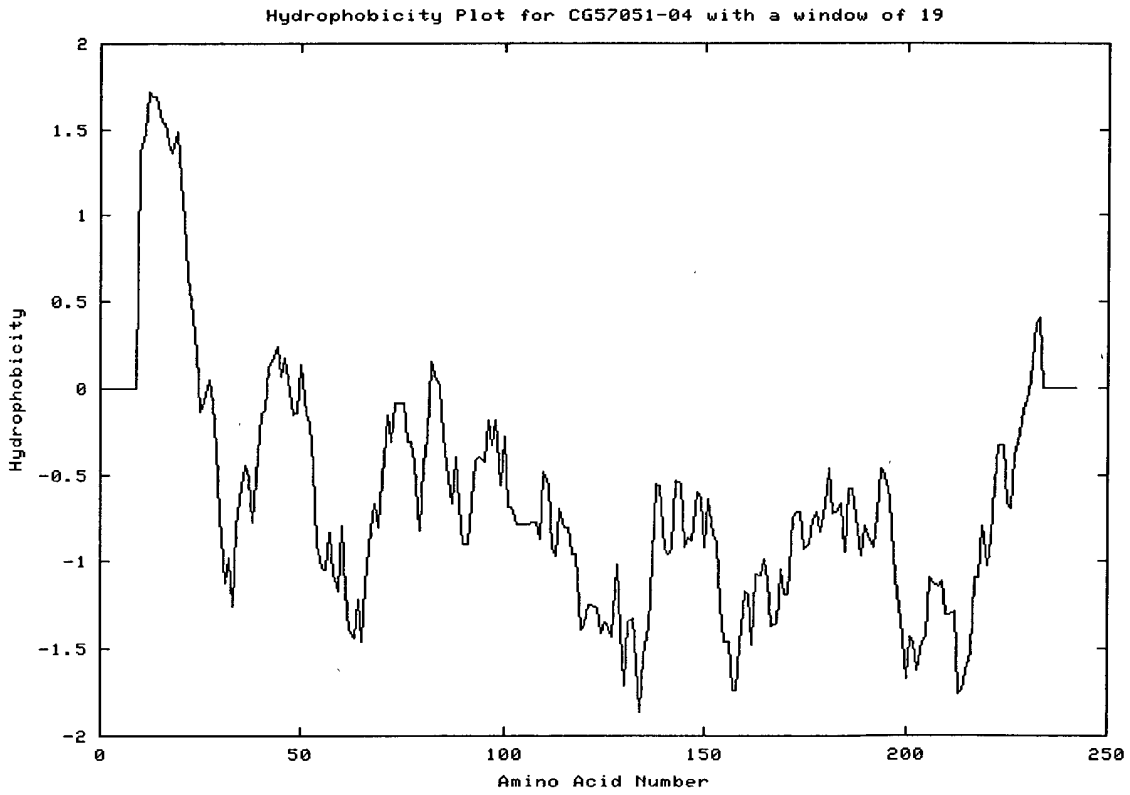
endoplasmic reticulum (membrane) --- Certainty=0.8200(Affirmative) < succ>
plasma membrane --- Certainty=0.1900(Affirmative) < succ>
microbody (peroxisome) --- Certainty=0.1701(Affirmative) < succ>
endoplasmic reticulum (lumen) --- Certainty=0.1000(Affirmative) < succ>

INTEGRAL Likelihood = -4.04 Transmembrane 7 - 23 (4 - 25)

Seems to be a Type Ib (Nexo Ccyt) membrane protein
Is the sequence a signal peptide?

# Measure	Position	Value	Cutoff	Conclusion
max. C	31	0.427	0.37	YES
max. Y	31	0.473	0.34	YES
max. S	8	0.952	0.88	YES
mean S	1-30	0.738	0.48	YES

Most likely cleavage site between pos. 30 and 31: VQS-KS



SECP 16

A SECP16 nucleic acid and polypeptide according to the invention were obtained by exon linking and include the nucleic acid sequence (SEQ ID NO:52) and encoded polypeptide sequence (SEQ ID NO:53) of clone CG57051-05 directed toward novel Angiopoietin-like proteins and nucleic acids encoding them. Figure 21 illustrates the nucleic acid sequence and amino acid sequences respectively. This clone includes a nucleotide sequence (SEQ ID NO:52) of 1239 bp. The nucleotide sequence includes an open reading frame (ORF) beginning with an ATG initiation codon at nucleotides 80-82 and ending with a TAG stop codon at nucleotides 1184-1186. Putative untranslated regions, if any, are found upstream from the initiation codon and downstream from the termination codon. The encoded protein having 368 amino acid residues is presented using the one-letter code in Figure 21. The protein encoded by clone CG57051-05 is predicted by the PSORT program to be located extracellularly with a certainty of 0.7332 and has a signal peptide (see Table 28 below). The PCR product derived by exon linking, covering the entire open reading frame, was cloned into the pCR2.1 vector from Invitrogen to provide clone 157544::CG50847-01.891637.M13 and clone 157544::CG50847-01.891637.O5.

Similarities

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 867 of 1064 bases (81%) identical to a gb:GENBANK-ID:AF202636|acc:AF202636.1 mRNA from Homo sapiens (Homo sapiens angiopoietin-like protein PP1158 mRNA, complete cds) (See Table 24). The full amino acid sequence of the protein of the invention was found to have 185 of 192 amino acid residues (96%) identical to, and 185 of 192 amino acid residues (96%) similar to, the 406 amino acid residue ptnr:SPTREMBL-ACC:Q9HBV4 protein from Homo sapiens (Human) (ANGIOPOIETIN-LIKE PROTEIN PP1158) (See Table 25).

A multiple sequence alignment is given in Table 27, with the protein of the invention being shown on the first line in a ClustalW analysis comparing the protein of the invention with related protein sequences. Please note this sequence represents a splice form of Angiopoietin, missing exon 4, as indicated in positions 183 to 221 and with SNPs: V156G, A157G, T266M.

The presence of identifiable domains in the protein disclosed herein was determined by searches versus domain databases such as Pfam, PROSITE, ProDom, Blocks or Prints and then identified by the Interpro domain accession number. Significant domains are summarized below:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
fibrinogen_C	1/2	184	246 ..	47	123 ..	98.2	4e-27
fibrinogen_C	2/2	288	362 ..	178	272 .]	67.0	3.4e-18

IPR002181; (Fibrinogen_C)

Fibrinogen, the principal protein of vertebrate blood clotting is an hexamer containing two sets of three different chains (alpha, beta, and gamma), linked to each other by disulfide bonds. The N-terminal sections of these three chains are evolutionary related and contain the cysteines that participate in the cross-linking of the chains. However, there is no similarity between the C-terminal part of the alpha chain and that of the beta and gamma chains. The C-terminal part of the beta and gamma chains forms a domain of about 270 amino-acid residues. As shown in the schematic representation this domain contains four conserved cysteines involved in two disulfide bonds.

```
xxxxCxxxxxxxxxxxxCxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxCxxxxxCxxxxxxxxxxx
                 |           |                          |     |
                 +-----------+                          +-----+
```

'C': conserved cysteine involved in a disulfide bond.

(SEQ ID NO:126)

Such a domain has been recently found in other proteins which are listed below:

1) Two sea cucumber fibrinogen-like proteins (FReP-A and FReP-B). These are proteins, of about 260 amino acids, which have a fibrinogen beta/gamma C-terminal domain.

2) In the C-terminus of Drosophila protein scabrous (gene sca). Scabrous is involved in the regulation of neurogenesis in Drosophila and may encode a lateral inhibitor of R8 cells differentiation.

3) In the C-terminus of a mammalian T-cell specific protein of unknown function.

4) In the C-terminus of a human protein of unknown function which is encoded on the opposite strand of the steroid 21-hydroxylase/complement component C4 gene locus.

The function of this domain is not yet known, but it has been suggested that it could be involved in protein-protein interactions.

This indicates that the sequence of the invention has properties similar to those of other proteins known to contain this/these domain(s) and similar to the properties of these domains.

Chromosomal information:

The Angiopoietin-like gene disclosed in this invention maps to chromosome 19p13.3. This assignment was made using mapping information associated with genomic clones, public genes and ESTs sharing sequence identity with the disclosed sequence and CuraGen Corporation's Electronic Northern bioinformatic tool.

Tissue expression

The Angiopoietin-like gene disclosed in this invention is expressed in at least the following tissues: Adipose, Liver, Placenta. Expression information was derived from the tissue sources of the sequences that were included in the derivation of the sequence of CuraGen Acc. No. CG57051-05.

Cellular Localization and Sorting

The PSORT, SignalP and hydropathy profile for the Angiopoietin-like protein are shown in Table 28. The results predict that this sequence has a signal peptide and is likely to be localized extracellularly with a certainty of 0.7332. The signal peptide is predicted by SignalP to be cleaved between amino acids 25 and 26: AQG-GP.

Functional Variants and Homologs

The novel nucleic acid of the invention encoding a Angiopoietin-like protein includes the nucleic acid whose sequence is provided in Figure 21, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Figure 21 while still encoding a protein that maintains its Angiopoietin-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to the sequence of CuraGen Acc. No. CG57051-05, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 19% of the bases may be so changed.

The novel protein of the invention includes the Angiopoietin-like protein whose sequence is provided in Figure 21. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Figure 21 while still encoding a protein that maintains its Angiopoietin-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 4% of the amino acid residues may be so changed.

Chimeric and Fusion Proteins

The present invention includes chimeric or fusion proteins of the Angiopoietin-like protein, in which the Angiopoietin-like protein of the present invention is joined to a second polypeptide or protein that is not substantially homologous to the present novel protein. The second polypeptide can be fused to either the amino-terminus or carboxyl-terminus of the present CG57051-05 polypeptide. In certain embodiments a third nonhomologous polypeptide or protein may also be fused to the novel Angiopoietin-like protein such that the second nonhomologous polypeptide or protein is joined at the amino terminus, and the third nonhomologous polypeptide or protein is joined at the carboxyl terminus, of the CG57051-05 polypeptide. Examples of

nonhomologous sequences that may be incorporated as either a second or third polypeptide or protein include glutathione S-transferase, a heterologous signal sequence fused at the amino terminus of the Angiopoietin-like protein, an immunoglobulin sequence or domain, a serum protein or domain thereof (such as a serum albumin), an antigenic epitope, and a specificity motif such as $(His)_6$.

The invention further includes nucleic acids encoding any of the chimeric or fusion proteins described in the preceding paragraph.

Antibodies

The invention further encompasses antibodies and antibody fragments, such as Fab, $(Fab)_2$ or single chain FV constructs, that bind immunospecifically to any of the proteins of the invention. Also encompassed within the invention are peptides and polypeptides comprising sequences having high binding affinity for any of the proteins of the invention, including such peptides and polypeptides that are fused to any carrier particle (or biologically expressed on the surface of a carrier) such as a bacteriophage particle.

Uses of the Compositions of the Invention

The protein similarity information, expression pattern, cellular localization, and map location for the protein and nucleic acid disclosed herein suggest that this Angiopoietin-like protein may have important structural and/or physiological functions characteristic of the Angiopoietin family. Therefore, the nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications and as a research tool. These include serving as a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed. These also include potential therapeutic applications such as the following: (i) a protein therapeutic, (ii) a small molecule drug target, (iii) an antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) a nucleic acid useful in gene therapy (gene delivery/gene ablation), (v) an agent promoting tissue regeneration *in vitro* and *in vivo*, and (vi) a biological defense weapon.

The nucleic acids and proteins of the invention have applications in the diagnosis and/or treatment of various diseases and disorders. For example, the compositions of the present invention will have efficacy for the treatment of patients suffering from: type II diabetes, obesity,

colon cancer, diabetes mellitus, insulin-resistant, with acanthosis nigricans and hypertension, 3-methylglutaconicaciduria, type III; Cone-rod retinal dystrophy-2; DNA ligase I deficiency; Glutaricaciduria, type IIB Liposarcoma; Myotonic dystrophy as well as other diseases, disorders and conditions.

These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in diagnostic and/or therapeutic methods.

Table 24. BLASTN search using CuraGen Acc. No. CG57051-05.

```
>gb:GENBANK-ID:AF202636|acc:AF202636.1 Homo sapiens angiopoietin-like protein
           PP1158 mRNA, complete cds - Homo sapiens, 1943 bp. (SEQ ID NO:87)
           Length = 1943

  Plus Strand HSPs:

 Score = 3105 (465.9 bits), Expect = 2.0e-134, P = 2.0e-134
 Identities = 867/1064 (81%), Positives = 867/1064 (81%), Strand = Plus /
Plus

Query:      4 CGTCTCCAGTCCTCGCACCTGGAACCCCAACGTCCCCGAGAGTCCCCGAATCCCCGCTCC 63
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:     97 CGTCTCCAGTCCTCGCACCTGGAACCCCAACGTCCCCGAGAGTCCCCGAATCCCCGCTCC 156

Query:     64 CAGGCTACCTAAGAGGATGAGCGGCGCTCCGACGGCCGGGGCAGCCCTGATGCTCTGCGC 123
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Sbjct:    157 CAGGCTACCTAAGAGGATGAGCGGTGCTCCGACGGCCGGGGCAGCCCTGATGCTCTGCGC 216

Query:    124 CGCCACCGCCGTGCTACTGAGCGCTCAGGGCGGACCCGTGCAGTCCAAGTCGCCGCGCTT 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    217 CGCCACCGCCGTGCTACTGAGCGCTCAGGGCGGACCCGTGCAGTCCAAGTCGCCGCGCTT 276

Query:    184 TGCGTCCTGGGACGAGATGAATGTCCTGGCGCACGGACTCCTGCAGCTCGGCCAGGGGCT 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    277 TGCGTCCTGGGACGAGATGAATGTCCTGGCGCACGGACTCCTGCAGCTCGGCCAGGGGCT 336

Query:    244 GCGCGAACACGCGGAGCGCACCCGCAGTCAGCTGAGCGCGCTGGAGCGGCGCCTGAGCGC 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    337 GCGCGAACACGCGGAGCGCACCCGCAGTCAGCTGAGCGCGCTGGAGCGGCGCCTGAGCGC 396

Query:    304 GTGCGGGTCCGCCTGTCAGGGAACCGAGGGGTCCACCGACCTCCCGTTAGCCCCTGAGAG 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    397 GTGCGGGTCCGCCTGTCAGGGAACCGAGGGGTCCACCGACCTCCCGTTAGCCCCTGAGAG 456

Query:    364 CCGGGTGGACCCTGAGGTCCTTCACAGCCTGCAGACACAACTCAAGGCTCAGAACAGCAG 423
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    457 CCGGGTGGACCCTGAGGTCCTTCACAGCCTGCAGACACAACTCAAGGCTCAGAACAGCAG 516

Query:    424 GATCCAGCAACTCTTCCACAAGGTGGCCCAGCAGCAGCGGCACCTGGAGAAGCAGCACCT 483
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    517 GATCCAGCAACTCTTCCACAAGGTGGCCCAGCAGCAGCGGCACCTGGAGAAGCAGCACCT 576

Query:    484 GCGAATTCAGCATCTGCAAAGCCAGTTTGGCCTCCTGGACCACAAGCACCTAGACCATGA 543
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    577 GCGAATTCAGCATCTGCAAAGCCAGTTTGGCCTCCTGGACCACAAGCACCTAGACCATGA 636

Query:    544 GGGTGGC-AAGCCTGCCCGAAGAAAGAGGCTGCCCGAGATGGCCCAGCCAGTTGACCCGG 602
              || |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    637 GG-TGGCCAAGCCTGCCCGAAGAAAGAGGCTGCCCGAGATGGCCCAGCCAGTTGACCCGG 695

Query:    603 CTCACAATGTCAGCCGCCTGCACCA--TGG--AGGC-TGGACAGTAA-T-TCAGAGGC-G 654
              ||||||||||||||||||||||||   ||   |||  | | ||| |  | |   |||  |
```

```
Sbjct:   696 CTCACAATGTCAGCCGCCTGCACCGGCTGCCCAGGGATTGCCAGGAGCTGTTCCAGGTTG 755

Query:   655 CCACGATGGCTCAGTGGACTTCAACCGGCCCTGGGA-AGCCTACAAGGCGGGGTTTGGGG 713
Sbjct:   756 GGGAGA-GGCAGAGTGGACTATTTGAAATCCAGCCTCAGGGGTCTCCGCCATTTTTGGTG 814

Query:   714 ATCCCCACGGCGAGTTCTGGCTGG-GTCTGGAGAAGGTGCATAGCATCATGGGGGACCGC 772
Sbjct:   815 AACTGCAAGATGACCTCAGA-TGGAGGCTGGACA-G-TA-ATT-CAG-A--GGCG-CCAC 865

Query:   773 AACAGCCGCCTGGCCGTGCAGCTGCGGGACTGGGATGGCAAC--GCCGAGTTGCTGCAGT 830
Sbjct:   866 GATGGCTCAGTGGACTT-CAAC--CGGCCCTGGGAAGCCTACAAGGCGGGGTT-TGGGGA 921

Query:   831 TCTCCGTG-C-AC--CTGGGTGGCGA-GGACACGGCCTATAGCCTG-CAGCTCACTGCAC 884
Sbjct:   922 TCCCCACGGCGAGTTCTGGCTGGGTCTGGAGAAGGTGCATAGCATCACGGGGGACCGCAA 981

Query:   885 CCGTGGCC-GGCCA-GCTGG-GCGCCACCACCGTCCCACCCAGCGGCCTCTCCGTACCCT 941
Sbjct:   982 CAGCCGCCTGGCCGTGCAGCTGCGGGACTGGGATGGCAAC--GCCGAGT-TGC-TGCAGT 1037

Query:   942 TCTCCACTTGGGACCAGGATCACGACCTCCGCAGGGACA-AGAACTGC-GCCAAGAGCCT 999
Sbjct:  1038 TCTCCG--TGC-ACCTGGGTGGCGAGGACA-C-GGCCTATAGC-CTGCAGCTCACTGCAC 1091

Query:  1000 CTCTGGAGGCTGGTG-GTTTGGCACCTGCAGCCATTCCAACCTCAACGGCCAGTACTTCC 1058
Sbjct:  1092 C-C--GTGGCCGGCCAGCTGGGCGCCACCA-CCGTCCCA-CC-CAGCGGCCTCTCCGTAC 1145

Query:  1059 GCTCCATCC 1067
Sbjct:  1146 CCTTC-TCC 1153
```

Score = 3048 (457.3 bits), Expect = 7.4e-132, P = 7.4e-132
Identities = 658/699 (94%), Positives = 658/699 (94%), Strand = Plus / Plus

```
Query:   541 TGAGG-GTGGCAAGCCTGCCCGAAGAAAGAGGCTGCCCGAGATGGCCCAGCCAGTTGACC 599
Sbjct:   754 TGGGGAGAGGCA-GAGTGGACTATTTGAAATCCAGCCTCAGGGGTCTCCGCCATTTTT-- 810

Query:   600 CGGCTCACAATGTCAGCCG-CCTGCACCATGGAGGCTGGACAGTAATTCAGAGGCGCCAC 658
Sbjct:   811 -GG-TGA-ACTGCAAGATGACCT-CAG-ATGGAGGCTGGACAGTAATTCAGAGGCGCCAC 865

Query:   659 GATGGCTCAGTGGACTTCAACCGGCCCTGGGAAGCCTACAAGGCGGGGTTTGGGGATCCC 718
Sbjct:   866 GATGGCTCAGTGGACTTCAACCGGCCCTGGGAAGCCTACAAGGCGGGGTTTGGGGATCCC 925

Query:   719 CACGGCGAGTTCTGGCTGGGTCTGGAGAAGGTGCATAGCATCATGGGGGACCGCAACAGC 778
Sbjct:   926 CACGGCGAGTTCTGGCTGGGTCTGGAGAAGGTGCATAGCATCACGGGGGACCGCAACAGC 985

Query:   779 CGCCTGGCCGTGCAGCTGCGGGACTGGGATGGCAACGCCGAGTTGCTGCAGTTCTCCGTG 838
Sbjct:   986 CGCCTGGCCGTGCAGCTGCGGGACTGGGATGGCAACGCCGAGTTGCTGCAGTTCTCCGTG 1045

Query:   839 CACCTGGGTGGCGAGGACACGGCCTATAGCCTGCAGCTCACTGCACCCGTGGCCGGCCAG 898
Sbjct:  1046 CACCTGGGTGGCGAGGACACGGCCTATAGCCTGCAGCTCACTGCACCCGTGGCCGGCCAG 1105

Query:   899 CTGGGCGCCACCACCGTCCCACCCAGCGGCCTCTCCGTACCCTTCTCCACTTGGGACCAG 958
Sbjct:  1106 CTGGGCGCCACCACCGTCCCACCCAGCGGCCTCTCCGTACCCTTCTCCACTTGGGACCAG 1165

Query:   959 GATCACGACCTCCGCAGGGACAAGAACTGCGCCAAGAGCCTCTCTGGAGGCTGGTGGTTT 1018
Sbjct:  1166 GATCACGACCTCCGCAGGGACAAGAACTGCGCCAAGAGCCTCTCTGGAGGCTGGTGGTTT 1225

Query:  1019 GGCACCTGCAGCCATTCCAACCTCAACGGCCAGTACTTCCGCTCCATCCCACAGCAGCGG 1078
Sbjct:  1226 GGCACCTGCAGCCATTCCAACCTCAACGGCCAGTACTTCCGCTCCATCCCACAGCAGCGG 1285

Query:  1079 CAGAAGCTTAAGAAGGGAATCTTCTGGAAGACCTGGCGGGGCCGCTACTACCCGCTGCAG 1138
Sbjct:  1286 CAGAAGCTTAAGAAGGGAATCTTCTGGAAGACCTGGCGGGGCCGCTACTACCCGCTGCAG 1345
```

```
Query:  1139 GCCACCACCATGTTGATCCAGCCCATGGCAGCAGAGGCAGCCTCCTAGCGTCCTGGCTGG 1198
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1346 GCCACCACCATGTTGATCCAGCCCATGGCAGCAGAGGCAGCCTCCTAGCGTCCTGGCTGG 1405

Query:  1199 GCCTGGTCCCAGGCCCACGAAAGA-GGTGACTCTTGGCTCTG 1239
             ||||||||||||||||||||||| ||||||||||||||||||
Sbjct:  1406 GCCTGGTCCCAGGCCCACGAAAGACGGTGACTCTTGGCTCTG 1447
```

Table 25. BLASTP search using the protein of CuraGen Acc. No. CG57051-05.

```
>ptnr:SPTREMBL-ACC:Q9HBV4 ANGIOPOIETIN-LIKE PROTEIN PP1158 - Homo sapiens
            (Human), 406 aa. (SEQ ID NO:88)
            Length = 406

 Score = 1015 (357.3 bits), Expect = 1.6e-197, Sum P(2) = 1.6e-197
 Identities = 185/192 (96%), Positives = 185/192 (96%)

Query:   177 NVSRLHHGGWTVIQRRHDGSVDFNRPWEAYKAGFGDPHGEFWLGLEKVHSIMGDRNSRLA 236
             |      ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Sbjct:   215 NCKMTSDGGWTVIQRRHDGSVDFNRPWEAYKAGFGDPHGEFWLGLEKVHSITGDRNSRLA 274

Query:   237 VQLRDWDGNAELLQFSVHLGGEDTAYSLQLTAPVAGQLGATTVPPSGLSVPFSTWDQDHD 296
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   275 VQLRDWDGNAELLQFSVHLGGEDTAYSLQLTAPVAGQLGATTVPPSGLSVPFSTWDQDHD 334

Query:   297 LRRDKNCAKSLSGGWWFGTCSHSNLNGQYFRSIPQQRQKLKKGIFWKTWRGRYYPLQATT 356
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   335 LRRDKNCAKSLSGGWWFGTCSHSNLNGQYFRSIPQQRQKLKKGIFWKTWRGRYYPLQATT 394

Query:   357 MLIQPMAAEAAS 368
             ||||||||||||
Sbjct:   395 MLIQPMAAEAAS 406
```

Score = 923 (324.9 bits), Expect = 1.6e-197, Sum P(2) = 1.6e-197
Identities = 180/182 (98%), Positives = 180/182 (98%)

```
Query:     1 MSGAPTAGAALMLCAATAVLLSAQGGPVQSKSPRFASWDEMNVLAHGLLQLGQGLREHAE 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:     1 MSGAPTAGAALMLCAATAVLLSAQGGPVQSKSPRFASWDEMNVLAHGLLQLGQGLREHAE 60

Query:    61 RTRSQLSALERRLSACGSACQGTEGSTDLPLAPESRVDPEVLHSLQTQLKAQNSRIQQLF 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    61 RTRSQLSALERRLSACGSACQGTEGSTDLPLAPESRVDPEVLHSLQTQLKAQNSRIQQLF 120

Query:   121 HKVAQQQRHLEKQHLRIQHLQSQFGLLDHKHLDHEGGKPARRKRLPEMAQPVDPAHNVSR 180
             ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||
Sbjct:   121 HKVAQQQRHLEKQHLRIQHLQSQFGLLDHKHLDHEVAKPARRKRLPEMAQPVDPAHNVSR 180

Query:   181 LH 182
             ||
Sbjct:   181 LH 182
```

Table 26. BLASTN identity search of CuraGen Corporation's Human SeqCalling database using CuraGen Acc. No. CG57051-05.

```
>s3aq:217939973 , 631 bp. (SEQ ID NO:89)
            Length = 631

  Minus Strand HSPs:

 Score = 2620 (393.1 bits), Expect = 9.1e-113, P = 9.1e-113
 Identities = 526/527 (99%), Positives = 526/527 (99%), Strand = Minus / Plus
```

```
Query:  1239 CAGAGCCAAGAGTCACC-TCTTTCGTGGGCCTGGGACCAGGCCCAGCCAGGACGCTAGGA 1181
             ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Sbjct:   105 CAGAGCCAAGAGTCACCGTCTTTCGTGGGCCTGGGACCAGGCCCAGCCAGGACGCTAGGA 164

Query:  1180 GGCTGCCTCTGCTGCCATGGGCTGGATCAACATGGTGGTGGCCTGCAGCGGGTAGTAGCG 1121
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   165 GGCTGCCTCTGCTGCCATGGGCTGGATCAACATGGTGGTGGCCTGCAGCGGGTAGTAGCG 224

Query:  1120 GCCCCGCCAGGTCTTCCAGAAGATTCCCTTCTTAAGCTTCTGCCGCTGCTGTGGGATGGA 1061
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   225 GCCCCGCCAGGTCTTCCAGAAGATTCCCTTCTTAAGCTTCTGCCGCTGCTGTGGGATGGA 284

Query:  1060 GCGGAAGTACTGGCCGTTGAGGTTGGAATGGCTGCAGGTGCCAAACCACCAGCCTCCAGA 1001
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   285 GCGGAAGTACTGGCCGTTGAGGTTGGAATGGCTGCAGGTGCCAAACCACCAGCCTCCAGA 344

Query:  1000 GAGGCTCTTGGCGCAGTTCTTGTCCCTGCGGAGGTCGTGATCCTGGTCCCAAGTGGAGAA 941
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   345 GAGGCTCTTGGCGCAGTTCTTGTCCCTGCGGAGGTCGTGATCCTGGTCCCAAGTGGAGAA 404

Query:   940 GGGTACGGAGAGGCCGCTGGGTGGACGGTGGTGGCGCCCAGCTGGCCGGCCACGGGTGC 881
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   405 GGGTACGGAGAGGCCGCTGGGTGGACGGTGGTGGCGCCCAGCTGGCCGGCCACGGGTGC 464

Query:   880 AGTGAGCTGCAGGCTATAGGCCGTGTCCTCGCCACCCAGGTGCACGGAGAACTGCAGCAA 821
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   465 AGTGAGCTGCAGGCTATAGGCCGTGTCCTCGCCACCCAGGTGCACGGAGAACTGCAGCAA 524

Query:   820 CTCGGCGTTGCCATCCCAGTCCCGCAGCTGCACGGCCAGGCGGCTGTTGCGGTCCCCCAT 761
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   525 CTCGGCGTTGCCATCCCAGTCCCGCAGCTGCACGGCCAGGCGGCTGTTGCGGTCCCCCAT 584

Query:   760 GATGCTATGCACCTTCTCCAGACCCAGCCAGAACTCGCCGTGGGGAT 714
             |||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   585 GATGCTATGCACCTTCTCCAGACCCAGCCAGAACTCGCCGTGGGGAT 631
```

```
>s3aq:230121563 , 788 bp. (SEQ ID NO:90)
          Length = 788

 Minus Strand HSPs:

Score = 2583 (387.6 bits), Expect = 3.4e-111, P = 3.4e-111
Identities = 533/548 (97%), Positives = 533/548 (97%), Strand = Minus / Plus

Query:  1239 CAGAGCCAAGAGTCACC-TCTTTCGTGGGCCTGGGACCAGGCCCAGCCAGGACGCTAGGA 1181
             ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Sbjct:   171 CAGAGCCAAGAGTCACCGTCTTTCGTGGGCCTGGGACCAGGCCCAGCCAGGACGCTAGGA 230

Query:  1180 GGCTGCCTCTGCTGCCATGGGCTGGATCAACATGGTGGTGGCCTGCAGCGGGTAGTAGCG 1121
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   231 GGCTGCCTCTGCTGCCATGGGCTGGATCAACATGGTGGTGGCCTGCAGCGGGTAGTAGCG 290

Query:  1120 GCCCCGCCAGGTCTTCCAGAAGATTCCCTTCTTAAGCTTCTGCCGCTGCTGTGGGATGGA 1061
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   291 GCCCCGCCAGGTCTTCCAGAAGATTCCCTTCTTAAGCTTCTGCCGCTGCTGTGGGATGGA 350

Query:  1060 GCGGAAGTACTGGCCGTTGAGGTTGGAATGGCTGCAGGTGCCAAACCACCAGCCTCCAGA 1001
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   351 GCGGAAGTACTGGCCGTTGAGGTTGGAATGGCTGCAGGTGCCAAACCACCAGCCTCCAGA 410

Query:  1000 GAGGCTCTTGGCGCAGTTCTTGTCCCTGCGGAGGTCGTGATCCTGGTCCCAAGTGGAGAA 941
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   411 GAGGCTCTTGGCGCAGTTCTTGTCCCTGCGGAGGTCGTGATCCTGGTCCCAAGTGGAGAA 470

Query:   940 GGGTACGGAGAGGCCGCTGGGTGGACGGTGGTGGCGCCCAGCTGGCCGGCCACGGGTGC 881
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   471 GGGTACGGAGAGGCCGCTGGGTGGACGGTGGTGGCGCCCAGCTGGCCGGCCACGGGTGC 530

Query:   880 AGTGAGCTGCAGGCTATAGGCCGTGTCCTCGCCACCCAGGTGCACGGAGAACTGCAGCAA 821
```

```
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  531 AGTGAGCTGCAGGCTATAGGCCGTGTCCTCGCCACCCAGGTGCACGGAGAACTGCAGCAA 590

Query:  820 CTCGGCGTTGCCATCCCAGTCCCGCAGCTGCACGGCCAGGCGGCTGTTGCGGTCCCCCAT 761
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Sbjct:  591 CTCGGCGTTGCCATCCCAGTCCCGCAGCTGCACGGCCAGGCGGCTGTTGCGGTCCCCCGT 650

Query:  760 GATGCTATGCACCTTCTCCAGACCCAGCCAGAACTCGCCGTGGGATCCCCAAACCCCGC 701
            ||||||||||||||||||||||||||||||||||||||| ||| |      |  || | |
Sbjct:  651 GATGCTATGCACCTTCTCCAGACCCAGCCAGAACTCGCC-TGGAGTGGGAGAGGCCACTC 709

Query:  700 CTTGTAGGC 692
            | || ||||
Sbjct:  710 CATG-AGGC 717
```

>s3aq:217940431 Category E: , 530 bp. (SEQ ID NO:91)

Length = 530

Minus Strand HSPs:

Score = 1795 (269.3 bits), Expect = 2.0e-75, P = 2.0e-75

Identities = 381/399 (95%), Positives = 381/399 (95%), Strand = Minus / Plus

```
Query:  553 CTTGCCACCCTCATGGTCTAGGTG-CTT-GTGGTCCAG-GAGGCCAAACTGGCTTTGCAG 497
            || | | || ||| | || | ||| ||  ||   |||  |||||||  ||||||||||||
Sbjct:  132 CTGGTCCCCGTCA-G-TCAATGTGACTGAGTCCGCCATTGAGGCCAGTCTGGCTTTGCAG 189

Query:  496 ATGCTGAATTCGCAGGTGCTGCTTCTCCAGGTGCCGCTGCTGCTGGGCCACCTTGTGGAA 437
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  190 ATGCTGAATTCGCAGGTGCTGCTTCTCCAGGTGCCGCTGCTGCTGGGCCACCTTGTGGAA 249

Query:  436 GAGTTGCTGGATCCTGCTGTTCTGAGCCTTGAGTTGTGTCTGCAGGCTGTGAAGGACCTC 377
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  250 GAGTTGCTGGATCCTGCTGTTCTGAGCCTTGAGTTGTGTCTGCAGGCTGTGAAGGACCTC 309

Query:  376 AGGGTCCACCCGGCTCTCAGGGGCTAACGGGAGGTCGGTGGACCCCTCGGTTCCCTGACA 317
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  310 AGGGTCCACCCGGCTCTCAGGGGCTAACGGGAGGTCGGTGGACCCCTCGGTTCCCTGACA 369

Query:  316 GGCGGACCCGCACGCGCTCAGGCGCCGCTCCAGCGCGCTCAGCTGACTGCGGGTGCGCTC 257
            ||||||||||||||||||||||||||| | ||||||||||||||||||||||||||||||
Sbjct:  370 GGCGGACCCGCACGCGCTCAGGCGCCGTTTCAGCGCGCTCAGCTGACTGCGGGTGCGCTC 429

Query:  256 CGCGTGTTCGCGCAGCCCCTGGCCGAGCTGCAGGAGTCCGTGCGCCAGGACATTCATCTC 197
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  430 CGCGTGTTCGCGCAGCCCCTGGCCGAGCTGCAGGAGTCCGTGCGCCAGGACATTCATCTC 489

Query:  196 GTCCCAGGACGCAAAGCGCGGCGACTTGGACTGCACGGGTC 156
            |||||||||||||||||||||||||||||||||||||||||
Sbjct:  490 GTCCCAGGACGCAAAGCGCGGCGACTTGGACTGCACGGGTC 530
```

>s3aq:217940613 , 336 bp. (SEQ ID NO:92)

Length = 336

Minus Strand HSPs:

Score = 995 (149.3 bits), Expect = 9.4e-56, Sum P(2) = 9.4e-56

Identities = 203/204 (99%), Positives = 203/204 (99%), Strand = Minus / Plus

```
Query:  626 GGTGCAGGCGGCTGACATTGTGAGCCGGGTCAACTGGCTGGGCCATCTCGGGCAGCCTCT 567
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  133 GGTGCAGGCGGCTGACATTGTGAGCCGGGTCAACTGGCTGGGCCATCTCGGGCAGCCTCT 192

Query:  566 TTCTTCGGCAGGCTTG-CCACCCTCATGGTCTAGGTGCTTGTGGTCCAGGAGGCCAAAC 508
            |||||||||||||||| ||||| |||||||||||||||||||||||||||||||||||||
Sbjct:  193 TTCTTCGGCAGGCTTGGCCACC-TCATGGTCTAGGTGCTTGTGGTCCAGGAGGCCAAAC 251

Query:  507 TGGCTTTGCAGATGCTGAATTCGCAGGTGCTGCTTCTCCAGGTGCCGCTGCTGCTGGGCC 448
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  252 TGGCTTTGCAGATGCTGAATTCGCAGGTGCTGCTTCTCCAGGTGCCGCTGCTGCTGGGCC 311
```

```
Query:   447 ACCTTGTGGAAGAGTTGCTGGATCC 423
             |||||||||||||||||||||||||
Sbjct:   312 ACCTTGTGGAAGAGTTGCTGGATCC 336

 Score = 410 (61.5 bits), Expect = 9.4e-56, Sum P(2) = 9.4e-56 (SEQ ID NO:129)
 Identities = 86/91 (94%), Positives = 86/91 (94%), Strand = Minus / Plus

Query:   717 GGATCCCCAAACCCCGCCTTGTAGGCTTCCCAGGGCCGGTTGAAGTCCACTGAGCCATCG 658
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:     1 GGATCCCCAAACCCCGCCTTGTAGGCTTCCCAGGGCCGGTTGAAGTCCACTGAGCCATCG 60

Query:   657 TGGCGCCTCTGAATTACTGTCCAGCCTCCAT 627
             |||||||||||||||| ||||||  || | |
Sbjct:    61 TGGCGCCTCTGAATTAATGTCCACTCTGCCT 91


>s3aq:217939964 , 328 bp. (SEQ ID NO:93)
           Length = 328

  Plus Strand HSPs:

 Score = 762 (114.3 bits), Expect = 1.5e-28, P = 1.5e-28
 Identities = 156/159 (98%), Positives = 156/159 (98%), Strand = Plus / Plus

Query:  1082 AAGCTTAAGAAGGGAATCTTCTGGAAGACCTGGCGGGGCCGCTACTACCCGCTGCAGGCC 1141
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:     1 AAGCTTAAGAAGGGAATCTTCTGGAAGACCTGGCGGGGCCGCTACTACCCGCTGCAGGCC 60

Query:  1142 ACCACCATGTTGATCCAGCCCATGGCAGCAGAGGCAGCCTCCTAGCGTCCTGGCTGGGCC 1201
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    61 ACCACCATGTTGATCCAGCCCATGGCAGCAGAGGCAGCCTCCTAGCGTCCTGGCTGGGCC 120

Query:  1202 TGGTCCCAGGCCCACGAAAGA-GGTGACTCTTGGCTCTG 1239
             |||||||||||| |||||||| |||||||||||||| |
Sbjct:   121 TGGTCCCAGGCCAACGAAAGACGGTGACTCTTGGCTCCG 159
```

Table 27. ClustalW alignment of CG57051-05 protein with related proteins.

```
CG57051-05   1  MSGAPTAGAALMLCAATAVLLSAQGGPVQSKSPRFASWDEMNVLAHGLLQLGQGLREHAE   60
Q9HBV4       1  MSGAPTAGAALMLCAATAVLLSAQGGPVQSKSPRFASWDEMNVLAHGLLQLGQGLREHAE   60
CG57051-04   1  MSGAPTAGAALMLCAATAVLLSARSGPVQSKSPRFASWDEMNVLAHGLLQLGQGLREHAE   60
CG57051-02   1  MSGAPTAGAALMLCAATAVLLSARSGPVQSKSPRFASWDEMNVLAHGLLQLGQGLREHAE   60

CG57051-05  61  RTRSQLSALERRLSACGSACQGTEGSTDLPLAPESRVDPEVLHSLQTQLKAQNSRIQQLF  120
Q9HBV4      61  RTRSQLSALERRLSACGSACQGTEGSTDLPLAPESRVDPEVLHSLQTQLKAQNSRIQQLF  120
CG57051-04  61  RTRSQLSALERRLSACGSACQGTEGSTDLPLAPESRVDPEVLHSLQTQLKAQNSRIQQLF  120
CG57051-02  61  RTRSQLSALERRLSACGSACQGTEGSTDLPLAPESRVDPEVLHSLQTQLKAQNSRIQQLF  120

CG57051-05 121  HKVAQQQRHLEKQHLRIQHLQSQFGLLDHKHLDHEGGKPARRKRLPEMAQPVDPAHNVSR  180
Q9HBV4     121  HKVAQQQRHLEKQHLRIQHLQSQFGLLDHKHLDHEVAKPARRKRLPEMAQPVDPAHNVSR  180
CG57051-04 121  HKVAQQQRHLEKQHLRIQHLQSQFGLLDHKHLDHEVAKPARRKRLPEMAQPVDPAHNVSR  180
CG57051-02 121  HKVAQQQRHLEKQHLRIQHLQSQFGLLDHKHLDHEVAKPARRKRLPEMAQPVDPAHNVSR  180

CG57051-05 181  LH--------------------------------------HGGWTVIQRRHDGSVDFNRP  202
Q9HBV4     181  LHRLPRDCQELFQVGERQSGLFEIQPQGSPPFLVNCKMTSDGGWTVIQRRHDGSVDFNRP  240
CG57051-04 181  LHR---------------------------------------------------------  183
CG57051-02 181  LHH--------------------------------------GGWTVIQRRHDGSMDFNRP  202

CG57051-05 203  WEAYKAGFGDPHGEFWLGLEKVHSIMGDRNSRLAVQLRDWDGNAELLQFSVHLGGEDTAY  262
Q9HBV4     241  WEAYKAGFGDPHGEFWLGLEKVHSITGDRNSRLAVQLRDWDGNAELLQFSVHLGGEDTAY  300
CG57051-04 183  ------------------------------------------------------------  183
CG57051-02 203  WEAYKAGFGDPHGEFWLGLEKVHSITGDRNSRLAVQLRDWDGNAELLQFSVHLGGEDTAY  262

CG57051-05 263  SLQLTAPVAGQLGATTVPPSGLSVPFSTWDQDHDLRRDKNCAKSL---------------  307
Q9HBV4     301  SLQLTAPVAGQLGATTVPPSGLSVPFSTWDQDHDLRRDKNCAKSL---------------  345
CG57051-04 183  ------------------------------------------------------------  183
CG57051-02 263  SLQLTAPVAGQLGATTVPPSGLSVPFSTWDQDHDLRRDKNCAKSLSAPSVAQRPDHVPSP  322

CG57051-05 308  ---SGGWWFGTCSHSNLNGQYFRSIPQQRQKLKKGIFWKTWRGRYYPLQATTMLIQPMAA  364
Q9HBV4     346  ---SGGWWFGTCSHSNLNGQYFRSIPQQRQKLKKGIFWKTWRGRYYPLQATTMLIQPMAA  402
CG57051-04 184  -----GWWFGTCSHSNLNGQYFRSIPQQRQKLKKGIFWKTWRGRYYPLQATTMLIQPMAA  238
CG57051-02 323  LTPAGGWWFGTCSHSNLNGQYFRSIPQQRQKLKKGIFWKTWRGRYYPLQATTMLIQPMAA  382

CG57051-05 365  EAAS                                                          368
Q9HBV4     403  EAAS                                                          406
CG57051-04 239  EAAS                                                          242
CG57051-02 383  EAAS                                                          386
```

Information for the ClustalW proteins:

Accno	Common Name	Length
CG57051-05 (SEQ ID NO:53)	novel Angiopoietin-like protein	368
CG57051-04 (SEQ ID NO:51)	Angiopoietin-like protein- isoform 4	242
CG57051-02 (SEQ ID NO:55)	Angiopoietin-like protein- isoform 2	386
Q9HBV4 (SEQ ID NO:80)	ANGIOPOIETIN-LIKE PROTEIN PP1158.	406

In the alignment shown above, black outlined amino acid residues indicate residues identically conserved between sequences (i.e., residues that may be required to preserve structural or functional properties); amino acid residues with a gray background are similar to one another between sequences, possessing comparable physical and/or chemical properties without altering protein structure or function (e.g. the group L,V, I, and M may be considered

similar); and amino acid residues with a white background are neither conserved nor similar between sequences.

Table 28. PSORT, SignalP and hydropathy results for CuraGen Acc. No. CG57051-05.

```
                    outside --- Certainty=0.7332(Affirmative) < succ>
          microbody (peroxisome) --- Certainty=0.2608(Affirmative) < succ>
endoplasmic reticulum (membrane) --- Certainty=0.1000(Affirmative) < succ>
  endoplasmic reticulum (lumen) --- Certainty=0.1000(Affirmative) < succ>

Is the sequence a signal peptide?
# Measure  Position  Value  Cutoff  Conclusion
 max. C   31      0.306  0.37   NO
 max. Y   26      0.429  0.34   YES
 max. S   8     0.952  0.88   YES
 mean S   1-25   0.848  0.48   YES
# Most likely cleavage site between pos. 25 and 26: AQG-GP
```

SECP 17

A SECP17 nucleic acid and polypeptide according to the invention includes the nucleic acid sequence (SEQ ID NO:54) and encoded polypeptide sequence (SEQ ID NO:55) of clone

CG57051-02 directed toward novel Angiopoietin-like proteins and nucleic acids encoding them. Figure 22 illustrates the nucleic acid sequence and amino acid sequences respectively. This clone includes a nucleotide sequence (SEQ ID NO:54) of 1315 bp. The nucleotide sequence includes an open reading frame (ORF) beginning with an ATG initiation codon at nucleotides 155-157 and ending with a TAG stop codon at nucleotides 1313-1315. Putative untranslated regions, if any, are found upstream from the initiation codon and downstream from the termination codon. The encoded protein having 386 amino acid residues is presented using the one-letter code in Figure 22. The protein encoded by clone CG57051-02 is predicted by the PSORT program to be located extracellularly with a certainty of 0.7332 and has a signal peptide (see Table 33 below). The PCR product derived by exon linking, covering the entire open reading frame, was cloned into the pCR2.1 vector from Invitrogen to provide clone 157544::CG50847-01.891637.M13 and clone 157544::CG50847-01.891637.O5. SeqCalling procedures were also utilized to identify CG57051-02, and the following public components were thus included in the invention: gb_accno: AC010323 Homo sapiens chromosome 19 clone CTD-2550O8, WORKING DRAFT SEQUENCE, 55 unordered pieces. In addition, the following Curagen Corporation SeqCalling Assembly ID's were also included in the invention:

162377751. The DNA and protein sequences for the novel Angiopoietin-like gene are reported here as CuraGen Acc. No. CG57051-02.

Similarities

CG57051-04 directed toward novel Angiopoietin-like proteins and nucleic acids encoding them. Figure 20 illustrates the nucleic acid sequence and amino acid sequences respectively. This clone includes a nucleotide sequence (SEQ ID NO:50) of 937 bp. The nucleotide sequence includes an open reading frame (ORF) beginning with an ATG initiation codon at nucleotides 155-157 and ending with a TAG stop codon at nucleotides 881-883. Putative untranslated regions, if any, are found upstream from the initiation codon and downstream from the termination codon. The encoded protein having 242 amino acid residues is presented using the one-letter code in Figure 20. The protein encoded by clone CG57051-04 is predicted by the PSORT program to be located at the endoplasmic reticulum with a certainty of 0.8200, and appears to be a signal protein (see Table 27 below).

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 696 of 700 bases (99%) identical to a gb:GENBANK-ID:AF202636|acc:AF202636.1 mRNA from Homo sapiens (Homo sapiens angiopoietin-like protein PP1158 mRNA, complete cds) (Table 29). The full amino acid sequence of the protein of the invention was found to have 179 of 182 amino acid residues (98%) identical to, and 180 of 182 amino acid residues (98%) similar to, the 406 amino acid residue ptnr:SPTREMBL-ACC:Q9NZU4 protein from Homo sapiens (Human) (HEPATIC ANGIOPOIETIN-RELATED PROTEIN) (Table 30).

A multiple sequence alignment is given in Table 32, with the protein of the invention being shown on the first line in a ClustalW analysis comparing the protein of the invention with related protein sequences.

The presence of identifiable domains in the protein disclosed herein was determined by searches versus domain databases such as Pfam, PROSITE, ProDom, Blocks or Prints and then identified by the Interpro domain accession number. Significant domains are summarized below:

```
hmmpfam - search a single seq against HMM database
HMMER 2.1.1 (Dec 1998)
Copyright (C) 1992-1998 Washington University School of Medicine
HMMER is freely distributed under the GNU General Public License (GPL).
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
HMM file:                 pfamHMMs
Sequence file.            /data4/genetools/kspytek39627Cg57051_02ProteinFasta.txt
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Query:  CG57051_02

Scores for sequence family classification (score includes all domains):
Model        Description                                   Score    E-value  N
--------     -----------                                   -----    ------- ---
fibrinogen_C Fibrinogen beta and gamma chains, C-term      143.9    3.6e-40   2

Parsed for domains:
Model        Domain  seq-f seq-t    hmm-f hmm-t       score  E-value
--------     ------- ----- -----    ----- -----       -----  -------
fibrinogen_C   1/2      184   246 ..     47   123 ..   102.5  2.4e-28
fibrinogen_C   2/2      288   380 ..    178   272 .]    43.4  1.9e-11

Alignments of top-scoring domains:
fibrinogen_C: domain 1 of 2, from 184 to 246: score 102.5, E = 2.4e-28
                   *->GGWTVfQrRqDGslnFyRnWkdYkeGFGnlstsgtGkkYCglpgEFW
                      GGWTV+QrR DGs +F+R W++Yk+GFG++              gEFW
  CG57051_02   184    GGWTVIQRRHDGSMDFNRPWEAYKAGFGDPH------------GEFW 218

                   LGNdkihlLTkqgsipyeLRveLeDwnGet<-*
                   LG++k h++T       + L v+L+Dw+G++
  CG57051_02   219 LGLEKVHSITGDR--NSRLAVQLRDWDGNA     245

fibrinogen_C: domain 2 of 2, from 288 to 380: score 43.4, E = 1.9e-11
                   *->FSTyDrDWDgWsTtspsgnCAesyg.....................gGRG
                      FST+D D D  +     ++nCA+s + ++  +++++ +++ ++ gG
  CG57051_02   288    FSTWDQDHD--L--RRDKNCAKSLSapsvaqrpdhvpspltpaGG-- 328

                   aWWynsChaAWLNGrYY....yGgtyspqEmaphGtDnGvvWatWkGsnq
                    WW+   C  +WLNG Y ++ +++ ++ +            G++W tW+G+
  CG57051_02   329 -WWFGTCSHSWLNGQYFrsipQQRQKLKK---------GIFWKTWRGR-- 365

                   AqPGGYwySmkfaeMKiRPr<-*
                         y ++ ++M i P
  CG57051_02   367 ------YYPLQATTMLIQPM    380
```

IPR002181: Fibrinogen [1] , the principal protein of vertebrate blood clotting is an hexamer containing two sets of three different chains (alpha, beta, and gamma), linked to each other by disulfide bonds. The N-terminal sections of these three chains are evolutionary related and contain the cysteines that participate in the cross-linking of the chains. However, there is no similarity between the C-terminal part of the alpha chain and that of the beta and gamma chains. The C-terminal part of the beta and gamma chains forms a domain of about 270 amino-acid residues. As shown in the schematic representation this domain contains four conserved cysteines involved in two disulfide bonds.

```
 xxxxCxxxxxxxxxxxxCxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxCxxxxxCxxxxxxxxxxxx
|              |                                                |      |
+--------------+                                                +------+
```

'C': conserved cysteine involved in a disulfide bond. (SEQ ID NO:126)

Such a domain has been recently found [2] in other proteins which are listed below.

Two sea cucumber fibrinogen-like proteins (FReP-A and FReP-B). These are proteins, of about 260 amino acids, which have a fibrinogen beta/gamma C-terminal domain. In the C-terminus of Drosophila protein scabrous (gene sca). Scabrous is involved in the regulation of neurogenesis in Drosophila and may encode a lateral inhibitor of R8 cells differentiation. In the C-terminus of a mammalian T-cell specific protein of unknown function. In the C-terminus of a human protein of unknown function which is encoded on the opposite strand of the steroid 21-hydroxylase/complement component C4 gene locus.

The function of this domain is not yet known, but it has been suggested [2] that it could be involved in protein-protein interactions.

This indicates that the sequence of the invention has properties similar to those of other proteins known to contain this/these domain(s) and similar to the properties of these domains.

Chromosomal information:

The Angiopoietin-like gene disclosed in this invention maps to chromosome 19q13.3. This assignment was made using mapping information associated with genomic clones, public genes and ESTs sharing sequence identity with the disclosed sequence and CuraGen Corporation's Electronic Northern bioinformatic tool.

Tissue expression

The Angiopoietin-like gene disclosed in this invention is expressed in at least the following tissues: adipocytes. Expression information was derived from the tissue sources of the sequences that were included in the derivation of the sequence of CuraGen Acc. No. CG57051-02.

Cellular Localization and Sorting

The PSORT, SignalP and hydropathy profile for the Angiopoietin-like protein are shown in Table 33. Although PSORT suggests that the Angiopoietin-like protein may be localized in the nucleus, the protein of CuraGen Acc. No. CG57051-02 predicted here is similar to the

Angiopoietin family, some members of which are secreted. Therefore it is likely that this novel Angiopoietin-like protein is localized to the same sub-cellular compartment.

Functional Variants and Homologs

The novel nucleic acid of the invention encoding an Angiopoietin-like protein includes the nucleic acid whose sequence is provided in Figure 22, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Figure 22 while still encoding a protein that maintains its Angiopoietin-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to the sequence of CuraGen Acc. No. CG57051-02, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 1% of the bases may be so changed.

The novel protein of the invention includes the Angiopoietin-like protein whose sequence is provided in Figure 22. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Figure 22 while still encoding a protein that maintains its Angiopoietin-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 2% of the amino acid residues may be so changed.

Antibodies

The invention further encompasses antibodies and antibody fragments, such as Fab, $(Fab)_2$ or single chain FV constructs, that bind immunospecifically to any of the proteins of the invention. Also encompassed within the invention are peptides and polypeptides comprising sequences having high binding affinity for any of the proteins of the invention, including such

peptides and polypeptides that are fused to any carrier particle (or biologically expressed on the surface of a carrier) such as a bacteriophage particle.

Uses of the Compositions of the Invention

The protein similarity information, expression pattern, cellular localization, and map location for the protein and nucleic acid disclosed herein suggest that this Angiopoietin-like protein may have important structural and/or physiological functions characteristic of the Angiopoietin family. Therefore, the nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications and as a research tool. These include serving as a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed. These also include potential therapeutic applications such as the following: (i) a protein therapeutic, (ii) a small molecule drug target, (iii) an antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) a nucleic acid useful in gene therapy (gene delivery/gene ablation), (v) an agent promoting tissue regeneration *in vitro* and *in vivo*, and (vi) a biological defense weapon.

The nucleic acids and proteins of the invention have applications in the diagnosis and/or treatment of various diseases and disorders. For example, the compositions of the present invention will have efficacy for the treatment of patients suffering from: type II diabetes, obesity, colon cancer, DIABETES MELLITUS, INSULIN-RESISTANT, WITH ACANTHOSIS NIGRICANS AND HYPERTENSION,3-methylglutaconicaciduria, type III; Cone-rod retinal dystrophy-2;DNA ligase I deficiency; Glutaricaciduria, type IIB;Liposarcoma; Myotonic dystrophy as well as other diseases, disorders and conditions.

These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in diagnostic and/or therapeutic methods.

Table 29. BLASTN search using CuraGen Acc. No. CG57051-02.

```
>gb:GENBANK-ID:AF202636|acc:AF202636.1 Homo sapiens angiopoietin-like protein
            PP1158 mRNA, complete cds - Homo sapiens, 1943 bp. (SEQ ID NO:94)
            Length = 1943

 Plus Strand HSPs:

Score = 3448 (517.3 bits), Expect = 8.3e-233, Sum P(2) = 8.3e-233
Identities = 696/700 (99%), Positives = 696/700 (99%), Strand = Plus / Plus
```

```
Query:     2 GCGGATCCTCACACGACTGTGATCCGATTCTTTCCAGCGGCTTCTGCAACCAAGCGGGTC 61
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    20 GCGGATCCTCACACGACTGTGATCCGATTCTTTCCAGCGGCTTCTGCAACCAAGCGGGTC 79

Query:    62 TTACCCCCGGTCCTCCGCGTCTCCAGTCCTCGCACCTGGAACCCCAACGTCCCCGAGAGT 121
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    80 TTACCCCCGGTCCTCCGCGTCTCCAGTCCTCGCACCTGGAACCCCAACGTCCCCGAGAGT 139

Query:   122 CCCCGAATCCCCGCTCCCAGGCTACCTAAGAGGATGAGCGGTGCTCCGACGGCCGGGGCA 181
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   140 CCCCGAATCCCCGCTCCCAGGCTACCTAAGAGGATGAGCGGTGCTCCGACGGCCGGGGCA 199

Query:   182 GCCCTGATGCTCTGCGCCGCCACCGCCGTGCTACTGAGCGCT-AGATCTGGACCCGTGCA 240
             |||||||||||||||||||||||||||||||||||||||||| ||  | |||||||||||
Sbjct:   200 GCCCTGATGCTCTGCGCCGCCACCGCCGTGCTACTGAGCGCTCAGGGC-GGACCCGTGCA 258

Query:   241 GTCCAAGTCGCCGCGCTTTGCGTCCTGGGACGAGATGAATGTCCTGGCGCACGGACTCCT 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   259 GTCCAAGTCGCCGCGCTTTGCGTCCTGGGACGAGATGAATGTCCTGGCGCACGGACTCCT 318

Query:   301 GCAGCTCGGCCAGGGGCTGCGCGAACACGCGGAGCGCACCCGCAGTCAGCTGAGCGCGCT 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   319 GCAGCTCGGCCAGGGGCTGCGCGAACACGCGGAGCGCACCCGCAGTCAGCTGAGCGCGCT 378

Query:   361 GGAGCGGCGCCTGAGCGCGTGCGGGTCCGCCTGTCAGGGAACCGAGGGGTCCACCGACCT 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   379 GGAGCGGCGCCTGAGCGCGTGCGGGTCCGCCTGTCAGGGAACCGAGGGGTCCACCGACCT 438

Query:   421 CCCGTTAGCCCCTGAGAGCCGGGTGGACCCTGAGGTCCTTCACAGCCTGCAGACACAACT 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   439 CCCGTTAGCCCCTGAGAGCCGGGTGGACCCTGAGGTCCTTCACAGCCTGCAGACACAACT 498

Query:   481 CAAGGCTCAGAACAGCAGGATCCAGCAACTCTTCCACAAGGTGGCCCAGCAGCAGCGGCA 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   499 CAAGGCTCAGAACAGCAGGATCCAGCAACTCTTCCACAAGGTGGCCCAGCAGCAGCGGCA 558

Query:   541 CCTGGAGAAGCAGCACCTGCGAATTCAGCATCTGCAAAGCCAGTTTGGCCTCCTGGACCA 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   559 CCTGGAGAAGCAGCACCTGCGAATTCAGCATCTGCAAAGCCAGTTTGGCCTCCTGGACCA 618

Query:   601 CAAGCACCTAGACCATGAGGTGGCCAAACCTGCCCGAAGAAAGAGGCTGCCCGAGATGGC 660
             ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Sbjct:   619 CAAGCACCTAGACCATGAGGTGGCCAAGCCTGCCCGAAGAAAGAGGCTGCCCGAGATGGC 678

Query:   661 CCAGCCAGTTGACCCGGCTCACAATGTCAGCCGCCTGCACC 701
             |||||||||||||||||||||||||||||||||||||||||
Sbjct:   679 CCAGCCAGTTGACCCGGCTCACAATGTCAGCCGCCTGCACC 719

 Score = 1887 (283.1 bits), Expect = 8.3e-233, Sum P(2) = 8.3e-233
 Identities = 399/415 (96%), Positives = 399/415 (96%), Strand = Plus / Plus

Query:   694 CCTGCACCATGGAGGCTGGACAGTAATTCAGAGGCGCCACGATGGCTCAATGGACTTCAA 753
             ||| ||  ||||||||||||||||||||||||||||||||||||||||| ||||||||||
Sbjct:   828 CCT-CAG-ATGGAGGCTGGACAGTAATTCAGAGGCGCCACGATGGCTCAGTGGACTTCAA 885

Query:   754 CCGGCCCTGGGAAGCCTACAAGGCGGGGTTTGGGGATCCCCACGGCGAGTTCTGGCTGGG 813
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   886 CCGGCCCTGGGAAGCCTACAAGGCGGGGTTTGGGGATCCCCACGGCGAGTTCTGGCTGGG 945

Query:   814 TCTGGAGAAGGTGCATAGCATCACGGGGGACCGCAACAGCCGCCTGGCCGTGCAGCTGCG 873
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   946 TCTGGAGAAGGTGCATAGCATCACGGGGGACCGCAACAGCCGCCTGGCCGTGCAGCTGCG 1005

Query:   874 GGACTGGGATGGCAACGCCGAGTTGCTGCAGTTCTCCGTGCACCTGGGTGGCGAGGACAC 933
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1006 GGACTGGGATGGCAACGCCGAGTTGCTGCAGTTCTCCGTGCACCTGGGTGGCGAGGACAC 1065

Query:   934 GGCCTATAGCCTGCAGCTCACTGCACCCGTGGCCGGCCAGCTGGGCGCCACCACCGTCCC 993
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   1066 GGCCTATAGCCTGCAGCTCACTGCACCCGTGGCCGGCCAGCTGGGCGCCACCACCGTCCC 1125

Query:    994 ACCCAGCGGCCTCTCCGTACCCTTCTCCACTTGGGACCAGGATCACGACCTCCGCAGGGA 1053
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   1126 ACCCAGCGGCCTCTCCGTACCCTTCTCCACTTGGGACCAGGATCACGACCTCCGCAGGGA 1185

Query:   1054 CAAGAACTGCGCCAAGAGCCTCTCTGCCCCATCGGTGGCTCAAAGACCTG-A-CCAT 1108
              ||||||||||||||||||||||||||      | ||||| |     ||||| | ||||
Sbjct:   1186 CAAGAACTGCGCCAAGAGCCTCTCTGGAGGCT-GGTGGTTTGGC-ACCTGCAGCCAT 1240

 Score = 936 (140.4 bits), Expect = 6.1e-190, Sum P(2) = 6.1e-190
 Identities = 312/407 (76%), Positives = 312/407 (76%), Strand = Plus / Plus

Query:    909 CCGTGCACCTGGGTGGCGAGGACACGGCCTATAGCCTGCAGCTCACTGCACCCGTGGCCG 968
              ||||||| ||| | | |  |||   |||  |  ||| | || |  |||||    |  |||
Sbjct:    993 CCGTGCAGCTGCGGGACTGGGAT--GGCA-AC-GCC-G-AGTTG-CTGCAGTTCT--CCG 1043

Query:    969 GCCAGCTGGGCGCC-ACCAC-CGTCCCAC--CCAGCGGCCTCTCCGTACCCTTCTCCACT 1024
                || ||||| | | |  || || || |    || || || || | | |||| |  ||
Sbjct:   1044 TGCACCTGGGTGGCGAGGACACGGCCTATAGCCTGCAGC-TCACTGCACCCGTGGCCGGC 1102

Query:   1025 TGGGACCAGGATC-ACGACC-TCCGCAGGGACAAGAACTGCGCCAAGAGCCTCTCTGCCC 1082
                |  |  ||  | || ||| ||| ||    |    |  || | ||   | ||| ||| ||
Sbjct:   1103 CAG--CTGGGCGCCACCACCGTCC-CACCCAGCGGC-CT-CTCCGT-ACCCT-TCT-CCA 1154

Query:   1083 CATCGGT---GGCTCAAAGACCTGACCATGTTCCCT--CTCC-CCT-GACCCCGGCAGGA 1135
              | | ||      || |||  |||||   ||| |  |       || | ||  || ||  | |||
Sbjct:   1155 CTTGGGACCAGGATCAC-GACCTCCGCAGGGACAAGAACTGCGCCAAGAGCCTCTCTGGA 1213

Query:   1136 GGCTGGTGGTTTGGCACCTGCAGCCATTCCAACCTCAACGGCCAGTACTTCCGCTCCATC 1195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   1214 GGCTGGTGGTTTGGCACCTGCAGCCATTCCAACCTCAACGGCCAGTACTTCCGCTCCATC 1273

Query:   1196 CCACAGCAGCGGCAGAAGCTTAAGAAGGGAATCTTCTGGAAGACCTGGCGGGGCCGCTAC 1255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   1274 CCACAGCAGCGGCAGAAGCTTAAGAAGGGAATCTTCTGGAAGACCTGGCGGGGCCGCTAC 1333

Query:   1256 TACCCGCTGCAGGCCACCACCATGTTGATCCAGCCCATGGCAGCAGAGGCAGCCTCCTAG 1315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   1334 TACCCGCTGCAGGCCACCACCATGTTGATCCAGCCCATGGCAGCAGAGGCAGCCTCCTAG 1393
```

Table 30. BLASTP search using the protein of CuraGen Acc. No. CG57051-02.

```
>ptnr:SPTREMBL-ACC:Q9NZU4 HEPATIC ANGIOPOIETIN-RELATED PROTEIN - Homo sapiens
            (Human), 406 aa. (SEQ ID NO:95)
            Length = 406

 Score = 919 (323.5 bits), Expect = 4.9e-194, Sum P(3) = 4.9e-194
 Identities = 179/182 (98%), Positives = 180/182 (98%)

Query:      1 MSGAPTAGAALMLCAATAVLLSARSGPVQSKSPRFASWDEMNVLAHGLLQLGQGLREHAE 60
              ||||||||||||||||||||||+ ||||||||||||||||||||||||||||||||||||
Sbjct:      1 MSGAPTAGAALMLCAATAVLLSAQGGPVQSKSPRFASWDEMNVLAHGLLQLGQGLREHAE 60

Query:     61 RTRSQLSALERRLSACGSACQGTEGSTDLPLAPESRVDPEVLHSLQTQLKAQNSRIQQLF 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:     61 RTRSQLSALERRLSACGSACQGTEGSTDLPLAPESRVDPEVLHSLQTQLKAQNSRIQQLF 120

Query:    121 HKVAQQQRHLEKQHLRIQHLQSQFGLLDHKHLDHEVAKPARRKRLPEMAQPVDPAHNVSR 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Sbjct:    121 HKVAQQQRHLEKQHLRIQHLQSQFGLLDHKHLDHEVAKPARRKRLPEMAQPVDPPHNVSR 180

Query:    181 LH 182
              LH
Sbjct:    181 LH 182
```

```
 Score = 670 (235.9 bits), Expect = 4.9e-194, Sum P(3) = 4.9e-194
 Identities = 123/132 (93%), Positives = 124/132 (93%)

Query:   177 NVSRLHHGGWTVIQRRHDGSMDFNRPWEAYKAGFGDPHGEFWLGLEKVHSITGDRNSRLA 236
             |      |||||||||||||+|||||||||||||||||||||||||||||| ||||||||
Sbjct:   215 NCKMTSDGGWTVIQRRHDGSVDFNRPWEAYKAGFGDPHGEFWLGLEKVHSIMGDRNSRLA 274

Query:   237 VQLRDWDGNAELLQFSVHLGGEDTAYSLQLTAPVAGQLGATTVPPSGLSVPFSTWDQDHD 296
             |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Sbjct:   275 VQLRDWDGNAELLQFSVHLGGEDTAYSLQFTAPVAGQLGATTVPPSGLSVPFSTWDQDHD 334

Query:   297 LRRDKNCAKSLS 308
             ||||||||||||
Sbjct:   335 LRRDKNCAKSLS 346

 Score = 331 (116.5 bits), Expect = 4.9e-194, Sum P(3) = 4.9e-194
 Identities = 59/61 (96%), Positives = 60/61 (98%)

Query:   326 AGGWWFGTCSHSNLNGQYFRSIPQQRQKLKKGIFWKTWRGRYYPLQATTMLIQPMAAEAA 385
             +|||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Sbjct:   346 SGGWWFGTCSHSNLNGQYFRSIPQQRQKLKKGIFWKTWRGRYYSLQATTMLIQPMAAEAA 405

Query:   386 S 386
             |
Sbjct:   406 S 406

 Score = 46 (16.2 bits), Expect = 5.9e-33, Sum P(2) = 5.9e-33
 Identities = 14/40 (35%), Positives = 19/40 (47%)

Query:   255 LGGEDTA-YSLQLTAPVAGQLGATTVPPSGLSVPFSTWDQ 293
             + |  ||  +| | |  |  | |   |    |  |++||+
Sbjct:     1 MSGAPTAGAALMLCAATAVLLSAQGGPVQSKSPRFASWDE 40

 Score = 45 (15.8 bits), Expect = 7.6e-33, Sum P(2) = 7.6e-33
 Identities = 13/40 (32%), Positives = 19/40 (47%)

Query:     1 MSGAPTAGAALMLCAATAVLLSARSGPVQSKSPRFASWDE 40
             + |  ||  +|   |  |  | | + |    |  |++||+
Sbjct:   293 LGGEDTA-YSLQFTAPVAGQLGATTVPPSGLSVPFSTWDQ 331
```

Table 31. BLASTN identity search of CuraGen Corporation's Human SeqCalling database using CuraGen Acc. No. CG57051-02.

```
>s3aq:162377751 Category D: , 1920 bp. (SEQ ID NO:96)
            Length = 1920

  Minus Strand HSPs:

 Score = 3448 (517.3 bits), Expect = 1.5e-233, Sum P(2) = 1.5e-233
 Identities = 696/700 (99%), Positives = 696/700 (99%), Strand = Minus / Plus

Query:   701 GGTGCAGGCGGCTGACATTGTGAGCCGGGTCAACTGGCTGGGCCATCTCGGGCAGCCTCT 642
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1221 GGTGCAGGCGGCTGACATTGTGAGCCGGGTCAACTGGCTGGGCCATCTCGGGCAGCCTCT 1280

Query:   641 TTCTTCGGGCAGGTTTGGCCACCTCATGGTCTAGGTGCTTGTGGTCCAGGAGGCCAAACT 582
             |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1281 TTCTTCGGGCAGGCTTGGCCACCTCATGGTCTAGGTGCTTGTGGTCCAGGAGGCCAAACT 1340

Query:   581 GGCTTTGCAGATGCTGAATTCGCAGGTGCTGCTTCTCCAGGTGCCGCTGCTGCTGGGCCA 522
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1341 GGCTTTGCAGATGCTGAATTCGCAGGTGCTGCTTCTCCAGGTGCCGCTGCTGCTGGGCCA 1400
```

```
Query:   521 CCTTGTGGAAGAGTTGCTGGATCCTGCTGTTCTGAGCCTTGAGTTGTGTCTGCAGGCTGT 462
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1401 CCTTGTGGAAGAGTTGCTGGATCCTGCTGTTCTGAGCCTTGAGTTGTGTCTGCAGGCTGT 1460

Query:   461 GAAGGACCTCAGGGTCCACCCGGCTCTCAGGGGCTAACGGGAGGTCGGTGGACCCCTCGG 402
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1461 GAAGGACCTCAGGGTCCACCCGGCTCTCAGGGGCTAACGGGAGGTCGGTGGACCCCTCGG 1520

Query:   401 TTCCCTGACAGGCGGACCCGCACGCGCTCAGGCGCCGCTCCAGCGCGCTCAGCTGACTGC 342
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1521 TTCCCTGACAGGCGGACCCGCACGCGCTCAGGCGCCGCTCCAGCGCGCTCAGCTGACTGC 1580

Query:   341 GGGTGCGCTCCGCGTGTTCGCGCAGCCCCTGGCCGAGCTGCAGGAGTCCGTGCGCCAGGA 282
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1581 GGGTGCGCTCCGCGTGTTCGCGCAGCCCCTGGCCGAGCTGCAGGAGTCCGTGCGCCAGGA 1640

Query:   281 CATTCATCTCGTCCAGGACGCAAAGCGCGGCGACTTGGACTGCACGGTCCAGATCT-A 223
             |||||||||||||||||||||||||||||||||||||||||||||||||||| |  || |
Sbjct:  1641 CATTCATCTCGTCCAGGACGCAAAGCGCGGCGACTTGGACTGCACGGTCC-GCCCTGA 1699

Query:   222 GCGCTCAGTAGCACGGCGGTGGCGGCGCAGAGCATCAGGGCTGCCCCGGCCGTCGGAGCA 163
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1700 GCGCTCAGTAGCACGGCGGTGGCGGCGCAGAGCATCAGGGCTGCCCCGGCCGTCGGAGCA 1759

Query:   162 CCGCTCATCCTCTTAGGTAGCCTGGGAGCGGGGATTCGGGGACTCTCGGGGACGTTGGGG 103
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1760 CCGCTCATCCTCTTAGGTAGCCTGGGAGCGGGGATTCGGGGACTCTCGGGGACGTTGGGG 1819

Query:   102 TTCCAGGTGCGAGGACTGGAGACGCGGAGGACCGGGGGTAAGACCCGCTTGGTTGCAGAA 43
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1820 TTCCAGGTGCGAGGACTGGAGACGCGGAGGACCGGGGGTAAGACCCGCTTGGTTGCAGAA 1879

Query:    42 GCCGCTGGAAAGAATCGGATCACAGTCGTGTGAGGATCCGC 2
             |||||||||||||||||||||||||||||||||||||||||
Sbjct:  1880 GCCGCTGGAAAGAATCGGATCACAGTCGTGTGAGGATCCGC 1920

 Score = 1887 (283.1 bits), Expect = 1.5e-233, Sum P(2) = 1.5e-233 (SEQ ID NO:130)
 Identities = 399/415 (96%), Positives = 399/415 (96%), Strand = Minus / Plus

Query:  1108 ATGG-T-CAGGTCTTTGAGCCACCGATGGGGCAGAGAGGCTCTTGGCGCAGTTCTTGTCC 1051
             |||| | |||||     | ||||| |    ||||||||||||||||||||||||||||||
Sbjct:   700 ATGGCTGCAGGTGCCAAA-CCACC-AGCCTCCAGAGAGGCTCTTGGCGCAGTTCTTGTCC 757

Query:  1050 CTGCGGAGGTCGTGATCCTGGTCCCAAGTGGAGAAGGGTACGGAGAGGCCGCTGGGTGGG 991
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   758 CTGCGGAGGTCGTGATCCTGGTCCCAAGTGGAGAAGGGTACGGAGAGGCCGCTGGGTGGG 817

Query:   990 ACGGTGGTGGCGCCCAGCTGGCCGGCCACGGGTGCAGTGAGCTGCAGGCTATAGGCCGTG 931
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   818 ACGGTGGTGGCGCCCAGCTGGCCGGCCACGGGTGCAGTGAGCTGCAGGCTATAGGCCGTG 877

Query:   930 TCCTCGCCACCCAGGTGCACGGAGAACTGCAGCAACTCGGCGTTGCCATCCCAGTCCCGC 871
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   878 TCCTCGCCACCCAGGTGCACGGAGAACTGCAGCAACTCGGCGTTGCCATCCCAGTCCCGC 937

Query:   870 AGCTGCACGGCCAGGCGGCTGTTGCGGTCCCCCGTGATGCTATGCACCTTCTCCAGACCC 811
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   938 AGCTGCACGGCCAGGCGGCTGTTGCGGTCCCCCGTGATGCTATGCACCTTCTCCAGACCC 997

Query:   810 AGCCAGAACTCGCCGTGGGGATCCCCAAACCCCGCCTTGTAGGCTTCCCAGGGCCGGTTG 751
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   998 AGCCAGAACTCGCCGTGGGGATCCCCAAACCCCGCCTTGTAGGCTTCCCAGGGCCGGTTG 1057

Query:   750 AAGTCCATTGAGCCATCGTGGCGCCTCTGAATTACTGTCCAGCCTCCATGGTGCAGG 694
             ||||||| |||||||||||||||||||||||||||||||||||||||||  || |||
Sbjct:  1058 AAGTCCACTGAGCCATCGTGGCGCCTCTGAATTACTGTCCAGCCTCCATC-TG-AGG 1112

 Score = 936 (140.4 bits), Expect = 1.1e-190, Sum P(2) = 1.1e-190 (SEQ ID NO:131)
```

```
Identities = 312/407 (76%), Positives = 312/407 (76%), Strand = Minus / Plus

Query: 1315 CTAGGAGGCTGCCTCTGCTGCCATGGGCTGGATCAACATGGTGGTGGCCTGCAGCGGGTA 1256
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  547 CTAGGAGGCTGCCTCTGCTGCCATGGGCTGGATCAACATGGTGGTGGCCTGCAGCGGGTA 606

Query: 1255 GTAGCGGCCCCGCCAGGTCTTCCAGAAGATTCCCTTCTTAAGCTTCTGCCGCTGCTGTGG 1196
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  607 GTAGCGGCCCCGCCAGGTCTTCCAGAAGATTCCCTTCTTAAGCTTCTGCCGCTGCTGTGG 666

Query: 1195 GATGGAGCGGAAGTACTGGCCGTTGAGGTTGGAATGGCTGCAGGTGCCAAACCACCAGCC 1136
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  667 GATGGAGCGGAAGTACTGGCCGTTGAGGTTGGAATGGCTGCAGGTGCCAAACCACCAGCC 726

Query: 1135 TCCTGCCGGGGTCAGGG-G-AGAGG--GAACATGGTCAGGTCTTTGAGCCA---CCGATG 1083
            ||| |   || ||  || | ||     |  | ||   ||||| | || ||    || | |
Sbjct:  727 TCCAGAGAGGCTCTTGGCGCAGTTCTTGTCCCTGCGGAGGTCGT-GATCCTGGTCCCAAG 785

Query: 1082 GGGCAGAGAGGCTCTTGGCGCAGTTCTTGTCCCTGCGGA-GGTCGTGAT-CCTGGTCCCA 1025
             || ||| ||| |   || | ||  |   |   || ||| ||| |||   ||  |  |
Sbjct:  786 TGG-AGA-AGGGTAC-GGAG-AGGCCGC-TGGGTG-GGACGGTGGTGGCGCCCAG--CTG 837

Query: 1024 AGTGGAGAAGGGTACGGAGAGGCCGCTGGGTG--GGACG-GTGGTGGCG-CCCAGCTGGC 969
               ||  | |||| | | ||| | || || |   || || ||  | ||  ||||| ||
Sbjct:  838 GCCGGCCACGGGTGCAGTGAG-CTGCAGGCTATAGGCCGTGTCCTCGCCACCCAGGTGCA 896

Query:  968 CGGCCACGGGTGCAGTGAGCTGCAGGCTATAGGCCGTGTCCTCGCCACCCAGGTGCACGG 909
            |||  |    |||||  | || | |||  |  ||| |  ||   || | ||| |||||||
Sbjct:  897 CGGAGAAC--TGCAGCAA-CT-C-GGCGTT--GCCATC-CCAGTCC-CGCAGCTGCACGG 947
```

Table 32. ClustalW alignment of CG57051-02 protein with related proteins.

```
CG57051_02  MSGAPTAGAALMLCAATAVLLSARSGPVQSKSPRFASWDEMNVLAHGLLQLGQGLREHAE
Q9NZU4      MSGAPTAGAALMLCAATAVLLSAQGGPVQSKSPRFASWDEMNVLAHGLLQLGQGLREHAE

CG57051_02  RTRSQLSALERRLSACGSACQGTEGSTDLPLAPESRVDPEVLHSLQTQLKAQNSRIQQLF
Q9NZU4      RTRSQLSALERRLSACGSACQGTEGSTDLPLAPESRVDPEVLHSLQTQLKAQNSRIQQLF

CG57051_02  HKVAQQQRHLEKQHLRIQHLQSQFGLLDHKHLDHEVAKPARRKRLPEMAQPVDPAHNVSR
Q9NZU4      HKVAQQQRHLEKQHLRIQHLQSQFGLLDHKHLDHEVAKPARRKRLPEMAQPVDPPHNVSR

CG57051_02  LH--------------------------------------HGGWTVIQRRHDGSMDFNRP
Q9NZU4      LHRLPRDCQELFQVGERQSGLFEIQPQGSPPFLVNCKMTSDGGWTVIQRRHDGSVDFNRP

CG57051_02  WEAYKAGFGDPHGEFWLGLEKVHSITGDRNSRLAVQLRDWDGNAELLQFSVHLGGEDTAY
Q9NZU4      WEAYKAGFGDPHGEFWLGLEKVHSIMGDRNSRLAVQLRDWDGNAELLQFSVHLGGEDTAY

CG57051_02  SLQLTAPVAGQLGATTVPPSGLSVPFSTWDQDHDLRRDKNCAKSLSAPSVAQRPDHVPSP
Q9NZU4      SLQFTAPVAGQLGATTVPPSGLSVPFSTWDQDHDLRRDKNCAKSLS--------------

CG57051_02  LTPAGGWWFGTCSHSNLNGQYFRSIPQQRQKLKKGIFWKTWRGRYYPLQATTMLIQPMAA
Q9NZU4      ----GGWWFGTCSHSNLNGQYFRSIPQQRQKLKKGIFWKTWRGRYYSLQATTMLIQPMAA

CG57051_02  EAAS
Q9NZU4      EAAS
```

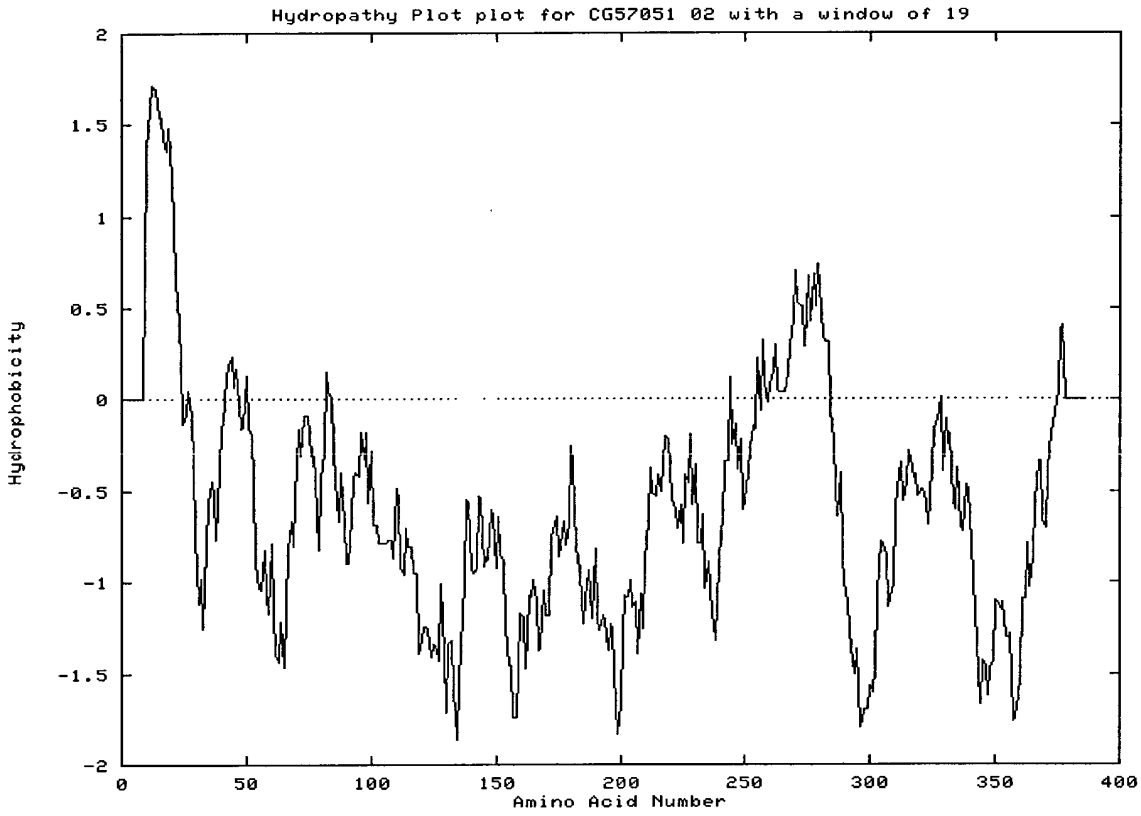
Information for the ClustalW proteins:

Accno	Common Name	Length
CG57051_02 (SEQ ID NO:55)	novel Angiopoietin-like protein	386
Q9NZU4 (SEQ ID NO:95)	HEPATIC ANGIOPOIETIN-RELATED PROTEIN.	406

In the alignment shown above, black outlined amino acid residues indicate residues identically conserved between sequences (i.e., residues that may be required to preserve structural or functional properties); amino acid residues with a gray background are similar to one another between sequences, possessing comparable physical and/or chemical properties without altering protein structure or function (e.g. the group L,V, I, and M may be considered similar); and amino acid residues with a white background are neither conserved nor similar between sequences.

Table 33. PSORT, SignalP and hydropathy results for CuraGen Acc. No. CG57051-02.

```
 endoplasmic reticulum (membrane) --- Certainty=0.8200(Affirmative) < succ>
           microbody (peroxisome) --- Certainty=0.3008(Affirmative) < succ>
                plasma membrane --- Certainty=0.1900(Affirmative) < succ>
   endoplasmic reticulum (lumen) --- Certainty=0.1000(Affirmative) < succ>

 INTEGRAL Likelihood = -4.04 Transmembrane 7 - 23 ( 4 - 25)

 Seems to be a Type Ib (Nexo Ccyt) membrane protein
Is the sequence a signal peptide?
# Measure  Position  Value  Cutoff  Conclusion
  max. C    31        0.427   0.37   YES
  max. Y    31        0.473   0.34   YES
  max. S     8        0.952   0.88   YES
  mean S     1-30     0.738   0.48   YES
# Most likely cleavage site between pos. 30 and 31: VQS-KS
```



SECP 18

A SECP18 nucleic acid and polypeptide according to the invention includes the nucleic acid sequence (SEQ ID NO:56) and encoded polypeptide sequence (SEQ ID NO:57) of clone

CG57051-03 directed toward novel Angiopoietin-like proteins and nucleic acids encoding them. Figure 23 illustrates the nucleic acid sequence and amino acid sequences respectively. This clone includes a nucleotide sequence (SEQ ID NO:56) of 1150 bp. The nucleotide sequence includes an open reading frame (ORF) beginning with an ATG initiation codon at nucleotides 44-46 and ending with a TAG stop codon at nucleotides 1148-1150. Putative untranslated regions, if any, are found upstream from the initiation codon and downstream from the termination codon. The encoded protein having 368 amino acid residues is presented using the one-letter code in Figure 23.

The protein encoded by clone CG57051-03 is predicted by the PSORT program to be located extracellularly with a certainty of 0.7332 and has a signal peptide (see Table 38 below).

The PCR product derived by exon linking, covering the entire open reading frame, was cloned into the pCR2.1 vector from Invitrogen to provide clone 134276::130294::PPAR-gamma.698782. P15. The DNA and protein sequences for the novel Angiopoietin-like gene are reported here as CuraGen Acc. No. CG57051-03.

Similarities

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 837 of 1031 bases (81%) identical to a gb:GENBANK-ID:AF202636|acc:AF202636.1 mRNA from Homo sapiens (Homo sapiens angiopoietin-like protein PP1158 mRNA, complete cds) (Table 34). The full amino acid sequence of the protein of the invention was found to have 184 of 192 amino acid residues (95%) identical to, and 184 of 192 amino acid residues (95%) similar to, the 406 amino acid residue ptnr:SPTREMBL-ACC:Q9HBV4 protein from Homo sapiens (Human) (ANGIOPOIETIN-LIKE PROTEIN PP1158) (Table 35).

A multiple sequence alignment is given in Table 37, with the protein of the invention being shown on the first line in a ClustalW analysis comparing the protein of the invention with related protein sequences. Please note this sequence represents a splice form of Angiopoietin as indicated in positions 183 to 221.

The presence of identifiable domains in the protein disclosed herein was determined by searches versus domain databases such as Pfam, PROSITE, ProDom, Blocks or Prints and then identified by the Interpro domain accession number. Significant domains are summarized below:

Model	Domain	seq-f	seq-t		hmm-f	hmm-t		score	E-value
fibrinogen_C	1/2	184	246	..	47	123	..	102.6	2.2e-28
fibrinogen_C	2/2	288	362	..	178	272	.]	61.3	1.4e-16

IPR002181; (Fibrinogen_C)

Fibrinogen, the principal protein of vertebrate blood clotting is an hexamer containing two sets of three different chains (alpha, beta, and gamma), linked to each other by disulfide bonds. The N-terminal sections of these three chains are evolutionary related and contain the cysteines that participate in the cross-linking of the chains. However, there is no similarity between the C-terminal part of the alpha chain and that of the beta and gamma chains. The C-

terminal part of the beta and gamma chains forms a domain of about 270 amino-acid residues. As shown in the schematic representation this domain contains four conserved cysteines involved in two disulfide bonds.

(SEQ ID NO:126)

```
xxxxCxxxxxxxxxxxxCxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxCxxxxxCxxxxxxxxxxx
                 |            |                                 |     |
                 +------------+                                 +-----+
```

'C': conserved cysteine involved in a disulfide bond.

Such a domain has been recently found in other proteins which are listed below:

1) Two sea cucumber fibrinogen-like proteins (FReP-A and FReP-B). These are proteins, of about 260 amino acids, which have a fibrinogen beta/gamma C-terminal domain.

2) In the C-terminus of Drosophila protein scabrous (gene sca). Scabrous is involved in the regulation of neurogenesis in Drosophila and may encode a lateral inhibitor of R8 cells differentiation.

3) In the C-terminus of a mammalian T-cell specific protein of unknown function.

4) In the C-terminus of a human protein of unknown function which is encoded on the opposite strand of the steroid 21-hydroxylase/complement component C4 gene locus.

The function of this domain is not yet known, but it has been suggested that it could be involved in protein-protein interactions.

This indicates that the sequence of the invention has properties similar to those of other proteins known to contain this/these domain(s) and similar to the properties of these domains.

Chromosomal information:

The Angiopoietin-like gene disclosed in this invention maps to chromosome 19p13.3. This assignment was made using mapping information associated with genomic clones, public genes and ESTs sharing sequence identity with the disclosed sequence and CuraGen Corporation's Electronic Northern bioinformatic tool.

Tissue expression

The Angiopoietin-like gene disclosed in this invention is expressed in at least the following tissues: Adipose, Liver, Placenta. Expression information was derived from the tissue sources of the sequences that were included in the derivation of the sequence of CuraGen Acc. No. CG57051-03.

Cellular Localization and Sorting

The PSORT, SignalP and hydropathy profile for the Angiopoietin-like protein are shown in Table 38. The results predict that this sequence has a signal peptide and is likely to be localized extracellularly with a certainty of 0.7332. The signal peptide is predicted by SignalP to be cleaved at amino acid 25 and 26: AQG-GP.

Functional Variants and Homologs

The novel nucleic acid of the invention encoding a Angiopoietin-like protein includes the nucleic acid whose sequence is provided in Figure 23, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Figure 23 while still encoding a protein that maintains its Angiopoietin-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to the sequence of CuraGen Acc. No. CG57051-03, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 19% of the bases may be so changed.

The novel protein of the invention includes the Angiopoietin-like protein whose sequence is provided in Figure 23. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Figure 23 while still encoding

a protein that maintains its Angiopoietin-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 5% of the amino acid residues may be so changed.

Chimeric and Fusion Proteins

The present invention includes chimeric or fusion proteins of the Angiopoietin-like protein, in which the Angiopoietin-like protein of the present invention is joined to a second polypeptide or protein that is not substantially homologous to the present novel protein. The second polypeptide can be fused to either the amino-terminus or carboxyl-terminus of the present CG57051-03 polypeptide. In certain embodiments a third nonhomologous polypeptide or protein may also be fused to the novel Angiopoietin-like protein such that the second nonhomologous polypeptide or protein is joined at the amino terminus, and the third nonhomologous polypeptide or protein is joined at the carboxyl terminus, of the CG57051-03 polypeptide. Examples of nonhomologous sequences that may be incorporated as either a second or third polypeptide or protein include glutathione S-transferase, a heterologous signal sequence fused at the amino terminus of the Angiopoietin-like protein, an immunoglobulin sequence or domain, a serum protein or domain thereof (such as a serum albumin), an antigenic epitope, and a specificity motif such as $(His)_6$.

The invention further includes nucleic acids encoding any of the chimeric or fusion proteins described in the preceding paragraph.

Antibodies

The invention further encompasses antibodies and antibody fragments, such as Fab, $(Fab)_2$ or single chain FV constructs, that bind immunospecifically to any of the proteins of the invention. Also encompassed within the invention are peptides and polypeptides comprising sequences having high binding affinity for any of the proteins of the invention, including such peptides and polypeptides that are fused to any carrier particle (or biologically expressed on the surface of a carrier) such as a bacteriophage particle.

Uses of the Compositions of the Invention

The protein similarity information, expression pattern, cellular localization, and map location for the protein and nucleic acid disclosed herein suggest that this Angiopoietin-like

protein may have important structural and/or physiological functions characteristic of the Fibrinogen family. Therefore, the nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications and as a research tool. These include serving as a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed. These also include potential therapeutic applications such as the following: (i) a protein therapeutic, (ii) a small molecule drug target, (iii) an antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) a nucleic acid useful in gene therapy (gene delivery/gene ablation), (v) an agent promoting tissue regeneration *in vitro* and *in vivo*, and (vi) a biological defense weapon.

The nucleic acids and proteins of the invention have applications in the diagnosis and/or treatment of various diseases and disorders. For example, the compositions of the present invention will have efficacy for the treatment of patients suffering from: type II diabetes, obesity, colon cancer, diabetes mellitus, insulin-resistant, with acanthosis nigricans and hypertension, 3-methylglutaconicaciduria, type III; Cone-rod retinal dystrophy-2;DNA ligase I deficiency; Glutaricaciduria, type IIB Liposarcoma; Myotonic dystrophy as well as other diseases, disorders and conditions.

These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in diagnostic and/or therapeutic methods.

Table 34. BLASTN search using CuraGen Acc. No. CG57051-03.

```
>gb:GENBANK-ID:AF202636|acc:AF202636.1 Homo sapiens angiopoietin-like protein
            PP1158 mRNA, complete cds - Homo sapiens, 1943 bp. (SEQ ID NO:97)
            Length = 1943

  Plus Strand HSPs:

 Score = 2967 (445.2 bits), Expect = 3.2e-128, P = 3.2e-128
 Identities = 837/1031 (81%), Positives = 837/1031 (81%), Strand = Plus / Plus

Query:      1 CCCCGAGAGTCCCCGAATCCCCGCTCCCAGGCTACCTAAGAGGATGAGCGGTGCTCCGAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    130 CCCCGAGAGTCCCCGAATCCCCGCTCCCAGGCTACCTAAGAGGATGAGCGGTGCTCCGAC 189

Query:     61 GGCCGGGGCAGCCCTGATGCTCTGCGCCGCCACCGCCGTGCTACTGAGCGCTCAGGGCGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    190 GGCCGGGGCAGCCCTGATGCTCTGCGCCGCCACCGCCGTGCTACTGAGCGCTCAGGGCGG 249

Query:    121 ACCCGTGCAGTCCAAGTCGCCGCGCTTTGCGTCCTGGGACGAGATGAATGTCCTGGCGCA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    250 ACCCGTGCAGTCCAAGTCGCCGCGCTTTGCGTCCTGGGACGAGATGAATGTCCTGGCGCA 309

Query:    181 CGGACTCCTGCAGCTCGGCCAGGGGCTGCGCGAACACGCGGAGCGCACCCGCAGTCAGCT 240
```

```
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  310 CGGACTCCTGCAGCTCGGCCAGGGGCTGCGCGAACACGCGGAGCGCACCCGCAGTCAGCT 369

Query:  241 GAGCGCGCTGGAGCGGCGCCTGAGCGCGTGCGGGTCCGCCTGTCAGGGAACCGAGGGGTC 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  370 GAGCGCGCTGGAGCGGCGCCTGAGCGCGTGCGGGTCCGCCTGTCAGGGAACCGAGGGGTC 429

Query:  301 CACCGACCTCCCGTTAGCCCCTGAGAGCCGGGTGGACCCTGAGGTCCTTCACAGCCTGCA 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  430 CACCGACCTCCCGTTAGCCCCTGAGAGCCGGGTGGACCCTGAGGTCCTTCACAGCCTGCA 489

Query:  361 GACACAACTCAAGGCTCAGAACAGCAGGATCCAGCAACTCTTCCACAAGGTGGCCCAGCA 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  490 GACACAACTCAAGGCTCAGAACAGCAGGATCCAGCAACTCTTCCACAAGGTGGCCCAGCA 549

Query:  421 GCAGCGGCACCTGGAGAAGCAGCACCTGCGAATTCAGCATCTGCAAAGCCAGTTTGGCCT 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  550 GCAGCGGCACCTGGAGAAGCAGCACCTGCGAATTCAGCATCTGCAAAGCCAGTTTGGCCT 609

Query:  481 CCTGGACCACAAGCACCTAGACCATGAGGTGGCCAAGCCTGCCCGAAGAAAGAGGCTGCC 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  610 CCTGGACCACAAGCACCTAGACCATGAGGTGGCCAAGCCTGCCCGAAGAAAGAGGCTGCC 669

Query:  541 CGAGATGGCCCAGCCAGTTGACCCGGCTCACAATGTCAGCCGCCTGCACCA--TGG--AG 596
            ||||||||||||||||||||||||||||||||||||||||||||||||||   ||   ||
Sbjct:  670 CGAGATGGCCCAGCCAGTTGACCCGGCTCACAATGTCAGCCGCCTGCACCGGCTGCCCAG 729

Query:  597 GC-TGGACAGTAA-T-TCAGAGGC-GCCACGATGGCTCAGTGGACTTCAACCGGCCCTGG 652
            |  | | ||| |  | |   |||  |    || |||  ||||||||         || |
Sbjct:  730 GGATTGCCAGGAGCTGTTCCAGGTTGGGGAGA-GGCAGAGTGGACTATTTGAAATCCAGC 788

Query:  653 GA-AGCCTACAAGGCGGGGTTTGGGGATCCCCACGGCGAGTTCTGGCTGG-GTCTGGAGA 710
               ||    |   ||    ||||| || |  || |  ||  || |  ||| | ||||| |
Sbjct:  789 CTCAGGGGTCTCCGCCATTTTTGGTGAACTGCAAGATGACCTCAGA-TGGAGGCTGGACA 847

Query:  711 AGGTCCATAGCATCACGGGGGACCGCAACAGCCGCCTGGCCGTGCAGCTGCGGGACTGGG 770
             | |  ||  ||  |  || | || | |  ||    ||| | | || |  |||  |||||
Sbjct:  848 -G-TA-ATT-CAG-A--GGCG-CCACGATGGCTCAGTGGACTT-CAAC--CGGCCCTGGG 896

Query:  771 ATG---ACAACGCCGAGTTGCTGCAGTTCTC-CGTGC-AC--CTGGGTGGCGA-GGACAC 822
            | |   |||| || | |||   | | | | | || || |   |||| |||    ||| |
Sbjct:  897 AAGCCTACAAGGCGGGGTTTGGGGA-TCCCCACG-GCGAGTTCTGGCTGGGTCTGGAGAA 954

Query:  823 GGCCTATAGCCTG-CAGCTCACTGCACCCGTGGCC-GGCCA-GCTGG-GCGCCACCACCG 878
            ||   ||||| |  | |   || ||| | |  ||| ||||  || |  |||  ||
Sbjct:  955 GGTGCATAGCATCACGGGGGACCGCAACAGCCGCCTGGCCGTGCAGCTGCGGGACTGGGA 1014

Query:  879 TCCCACCCAGCGGCCTCTCCGTACCCTTCCCCACTTGGGACCAGGATCACGACCTCCGCA 938
            |  || |  || |  | | | | |  ||| ||   ||  ||| || |  |||   |  |
Sbjct: 1015 TGGCAAC--GCCGAGT-TGC-TGCAGTTCTCCG--TGC-ACCTGGGTGGCGAGGACA-C- 1065

Query:  939 GGGACA-AGAACTGC-GCCAAGAGCCTCTCTGGAGGCTGGTG-GTTTGGCACCTGCAGCC 995
            ||   | ||  |||| ||  |  ||  | |  | ||| ||   | | ||| ||  || ||
Sbjct: 1066 GGCCTATAGC-CTGCAGCTCACTGCACC-C--GTGGCCGGCCAGCTGGGCGCCACCA-CC 1120

Query:  996 ATTCCAACCTCAACGGCCAGTACTTCCGCTCCATCC 1031
             | ||| || || |||||  | | | | || | |||
Sbjct: 1121 GTCCCA-CC-CAGCGGCCTCTCCGTACCCTTC-TCC 1153

 Score = 2774 (416.2 bits), Expect = 1.6e-119, P = 1.6e-119
 Identities = 562/568 (98%), Positives = 562/568 (98%), Strand = Plus / Plus

Query:  583 CCTGCACCATGGAGGCTGGACAGTAATTCAGAGGCGCCACGATGGCTCAGTGGACTTCAA 642
            ||| ||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  828 CCT-CAG-ATGGAGGCTGGACAGTAATTCAGAGGCGCCACGATGGCTCAGTGGACTTCAA 885

Query:  643 CCGGCCCTGGGAAGCCTACAAGGCGGGGTTTGGGGATCCCCACGGCGAGTTCTGGCTGGG 702
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Sbjct:   886 CCGGCCCTGGGAAGCCTACAAGGCGGGGTTTGGGGATCCCCACGGCGAGTTCTGGCTGGG 945

Query:   703 TCTGGAGAAGGTCCATAGCATCACGGGGGACCGCAACAGCCGCCTGGCCGTGCAGCTGCG 762
             |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   946 TCTGGAGAAGGTGCATAGCATCACGGGGGACCGCAACAGCCGCCTGGCCGTGCAGCTGCG 1005

Query:   763 GGACTGGGATGACAACGCCGAGTTGCTGCAGTTCTCCGTGCACCTGGGTGGCGAGGACAC 822
             ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1006 GGACTGGGATGGCAACGCCGAGTTGCTGCAGTTCTCCGTGCACCTGGGTGGCGAGGACAC 1065

Query:   823 GGCCTATAGCCTGCAGCTCACTGCACCCGTGGCCGGCCAGCTGGGCGCCACCACCGTCCC 882
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1066 GGCCTATAGCCTGCAGCTCACTGCACCCGTGGCCGGCCAGCTGGGCGCCACCACCGTCCC 1125

Query:   883 ACCCAGCGGCCTCTCCGTACCCTTCCCCACTTGGGACCAGGATCACGACCTCCGCAGGGA 942
             ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Sbjct:  1126 ACCCAGCGGCCTCTCCGTACCCTTCTCCACTTGGGACCAGGATCACGACCTCCGCAGGGA 1185

Query:   943 CAAGAACTGCGCCAAGAGCCTCTCTGGAGGCTGGTGGTTTGGCACCTGCAGCCATTCCAA 1002
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1186 CAAGAACTGCGCCAAGAGCCTCTCTGGAGGCTGGTGGTTTGGCACCTGCAGCCATTCCAA 1245

Query:  1003 CCTCAACGGCCAGTACTTCCGCTCCATCCCACAGCAGCGGCAGAAGCTTAAGAAGGGAAT 1062
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1246 CCTCAACGGCCAGTACTTCCGCTCCATCCCACAGCAGCGGCAGAAGCTTAAGAAGGGAAT 1305

Query:  1063 CTTCTGGAAGACCTGGCGGGGCCGCTACTACCCGCTGCAGGCCACCACCATGTTGATCCA 1122
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1306 CTTCTGGAAGACCTGGCGGGGCCGCTACTACCCGCTGCAGGCCACCACCATGTTGATCCA 1365

Query:  1123 GCCCATGGCAGCAGAGGCAGCCTCCTAG 1150
             ||||||||||||||||||||||||||||
Sbjct:  1366 GCCCATGGCAGCAGAGGCAGCCTCCTAG 1393
```

Table 35. BLASTP search using the protein of CuraGen Acc. No. CG57051-03.

```
>ptnr:SPTREMBL-ACC:Q9HBV4 ANGIOPOIETIN-LIKE PROTEIN PP1158 - Homo sapiens
          (Human), 406 aa. (SEQ ID NO:98)
          Length = 406

 Score = 1009 (355.2 bits), Expect = 4.3e-198, Sum P(2) = 4.3e-198
 Identities = 184/192 (95%), Positives = 184/192 (95%)

Query:   177 NVSRLHHGGWTVIQRRHDGSVDFNRPWEAYKAGFGDPHGEFWLGLEKVHSITGDRNSRLA 236
             |      |||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   215 NCKMTSDGGWTVIQRRHDGSVDFNRPWEAYKAGFGDPHGEFWLGLEKVHSITGDRNSRLA 274

Query:   237 VQLRDWDDNAELLQFSVHLGGEDTAYSLQLTAPVAGQLGATTVPPSGLSVPFPTWDQDHD 296
             ||||||| |||||||||||||||||||||||||||||||||||||||||||| |||||||
Sbjct:   275 VQLRDWDGNAELLQFSVHLGGEDTAYSLQLTAPVAGQLGATTVPPSGLSVPFSTWDQDHD 334

Query:   297 LRRDKNCAKSLSGGWWFGTCSHSNLNGQYFRSIPQQRQKLKKGIFWKTWRGRYYPLQATT 356
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   335 LRRDKNCAKSLSGGWWFGTCSHSNLNGQYFRSIPQQRQKLKKGIFWKTWRGRYYPLQATT 394

Query:   357 MLIQPMAAEAAS 368
             ||||||||||||
Sbjct:   395 MLIQPMAAEAAS 406

 Score = 934 (328.8 bits), Expect = 4.3e-198, Sum P(2) = 4.3e-198
 Identities = 182/182 (100%), Positives = 182/182 (100%)

Query:     1 MSGAPTAGAALMLCAATAVLLSAQGGPVQSKSPRFASWDEMNVLAHGLLQLGQGLREHAE 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Sbjct:     1 MSGAPTAGAALMLCAATAVLLSAQGGPVQSKSPRFASWDEMNVLAHGLLQLGQGLREHAE 60

Query:    61 RTRSQLSALERRLSACGSACQGTEGSTDLPLAPESRVDPEVLHSLQTQLKAQNSRIQQLF 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    61 RTRSQLSALERRLSACGSACQGTEGSTDLPLAPESRVDPEVLHSLQTQLKAQNSRIQQLF 120

Query:   121 HKVAQQQRHLEKQHLRIQHLQSQFGLLDHKHLDHEVAKPARRKRLPEMAQPVDPAHNVSR 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   121 HKVAQQQRHLEKQHLRIQHLQSQFGLLDHKHLDHEVAKPARRKRLPEMAQPVDPAHNVSR 180

Query:   181 LH 182
             ||
Sbjct:   181 LH 182
```

Table 36. BLASTN identity search of CuraGen Corporation's Human SeqCalling database using CuraGen Acc. No. CG57051-03.

```
>s3aq:189266374 Sequence 5 from Patent WO0105825 (AX079971.1: 100%/409,
            p=1.2e-238), 550 bp. (SEQ ID NO:99)
            Length = 550

  Plus Strand HSPs:

 Score = 2723 (408.6 bits), Expect = 1.8e-117, P = 1.8e-117
 Identities = 547/550 (99%), Positives = 547/550 (99%), Strand = Plus / Plus

Query:   450 GAATTCAGCATCTGCAAAGCCAGTTTGGCCTCCTGGACCACAAGCACCTAGACCATGAGG 509
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:     1 GAATTCAGCATCTGCAAAGCCAGTTTGGCCTCCTGGACCACAAGCACCTAGACCATGAGG 60

Query:   510 TGGCCAAGCCTGCCCGAAGAAAGAGGCTGCCCGAGATGGCCCAGCCAGTTGACCCGGCTC 569
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    61 TGGCCAAGCCTGCCCGAAGAAAGAGGCTGCCCGAGATGGCCCAGCCAGTTGACCCGGCTC 120

Query:   570 ACAATGTCAGCCGCCTGCACCATGGAGGCTGGACAGTAATTCAGAGGCGCCACGATGGCT 629
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   121 ACAATGTCAGCCGCCTGCACCATGGAGGCTGGACAGTAATTCAGAGGCGCCACGATGGCT 180

Query:   630 CAGTGGACTTCAACCGGCCCTGGGAAGCCTACAAGGCGGGGTTTGGGGATCCCCACGGCG 689
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   181 CAGTGGACTTCAACCGGCCCTGGGAAGCCTACAAGGCGGGGTTTGGGGATCCCCACGGCG 240

Query:   690 AGTTCTGGCTGGGTCTGGAGAAGGTCCATAGCATCACGGGGGACCGCAACAGCCGCCTGG 749
             ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Sbjct:   241 AGTTCTGGCTGGGTCTGGAGAAGGTGCATAGCATCACGGGGGACCGCAACAGCCGCCTGG 300

Query:   750 CCGTGCAGCTGCGGGACTGGGATGACAACGCCGAGTTGCTGCAGTTCTCCGTGCACCTGG 809
             ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Sbjct:   301 CCGTGCAGCTGCGGGACTGGGATGGCAACGCCGAGTTGCTGCAGTTCTCCGTGCACCTGG 360

Query:   810 GTGGCGAGGACACGGCCTATAGCCTGCAGCTCACTGCACCCGTGGCCGGCCAGCTGGGCG 869
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   361 GTGGCGAGGACACGGCCTATAGCCTGCAGCTCACTGCACCCGTGGCCGGCCAGCTGGGCG 420

Query:   870 CCACCACCGTCCCACCCAGCGGCCTCTCCGTACCCTTCCCCACTTGGGACCAGGATCACG 929
             |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Sbjct:   421 CCACCACCGTCCCACCCAGCGGCCTCTCCGTACCCTTCTCCACTTGGGACCAGGATCACG 480

Query:   930 ACCTCCGCAGGGACAAGAACTGCGCCAAGAGCCTCTCTGGAGGCTGGTGGTTTGGCACCT 989
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   481 ACCTCCGCAGGGACAAGAACTGCGCCAAGAGCCTCTCTGGAGGCTGGTGGTTTGGCACCT 540

Query:   990 GCAGCCATTC 999
```

```
                 ||||||||||
Sbjct:    541 GCAGCCATTC 550
```

>s3aq:188990257 Homo sapiens angiopoietin-related protein mRNA, complete cds
(AF153606.1: 99%/476, p=1.9e-259), 652 bp. (SEQ ID NO:100)
Length = 652

Minus Strand HSPs:

Score = 2403 (360.5 bits), Expect = 4.2e-103, P = 4.2e-103
Identities = 505/523 (96%), Positives = 505/523 (96%), Strand = Minus / Plus

```
Query:    520 AGGCTTGGCCACC-TCATGGTCTAGGTG-CTT-GTGGTCCAG-GAGGCCAAACTGGCTTT 465
              || | ||| | || ||| | || | ||| ||  ||   |||  |||||||  ||||||||
Sbjct:    128 AGCCCTGGTCCCCGTCA-G-TCAATGTGACTGAGTCCGCCATTGAGGCCAGTCTGGCTTT 185

Query:    464 GCAGATGCTGAATTCGCAGGTGCTGCTTCTCCAGGTGCCGCTGCTGCTGGGCCACCTTGT 405
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    186 GCAGATGCTGAATTCGCAGGTGCTGCTTCTCCAGGTGCCGCTGCTGCTGGGCCACCTTGT 245

Query:    404 GGAAGAGTTGCTGGATCCTGCTGTTCTGAGCCTTGAGTTGTGTCTGCAGGCTGTGAAGGA 345
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    246 GGAAGAGTTGCTGGATCCTGCTGTTCTGAGCCTTGAGTTGTGTCTGCAGGCTGTGAAGGA 305

Query:    344 CCTCAGGGTCCACCCGGCTCTCAGGGGCTAACGGGAGGTCGGTGGACCCCTCGGTTCCCT 285
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    306 CCTCAGGGTCCACCCGGCTCTCAGGGGCTAACGGGAGGTCGGTGGACCCCTCGGTTCCCT 365

Query:    284 GACAGGCGGACCCGCACGCGCTCAGGCGCCGCTCCAGCGCGCTCAGCTGACTGCGGGTGC 225
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    366 GACAGGCGGACCCGCACGCGCTCAGGCGCCGCTCCAGCGCGCTCAGCTGACTGCGGGTGC 425

Query:    224 GCTCCGCGTGTTCGCGCAGCCCCTGGCCGAGCTGCAGGAGTCCGTGCGCCAGGACATTCA 165
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    426 GCTCCGCGTGTTCGCGCAGCCCCTGGCCGAGCTGCAGGAGTCCGTGCGCCAGGACATTCA 485

Query:    164 TCTCGTCCCAGGACGCAAAGCGCGGCGACTTGGACTGCACGGGTCCGCCCTGAGCGCTCA 105
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    486 TCTCGTCCCAGGACGCAAAGCGCGGCGACTTGGACTGCACGGGTCCGCCCTGAGCGCTCA 545

Query:    104 GTAGCACGGCGGTGGCGGCGCAGAGCATCAGGGCTGCCCCGGCCGTCGGAGCACCGCTCA 45
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    546 GTAGCACGGCGGTGGCGGCGCAGAGCATCAGGGCTGCCCCGGCCGTCGGAGCACCGCTCA 605

Query:     44 TCCTCTTAGGTAGCCTGGGAGCGGGGATTCGGGGACTCT-CGGGG 1
              ||||||||||||||||||||||||||||||||||||||| |||||
Sbjct:    606 TCCTCTTAGGTAGCCTGGGAGCGGGGATTCGGGGACTCTTCGGGG 650
```

>s3aq:164987939 Category E: Homo sapiens angiopoietin-related protein mRNA,
complete cds (AF153606.1: 100%/150, p=1.9e-084), 228 bp. (SEQ ID NO:101)
Length = 228

Minus Strand HSPs:

Score = 480 (72.0 bits), Expect = 2.7e-31, Sum P(2) = 2.7e-31
Identities = 96/96 (100%), Positives = 96/96 (100%), Strand = Minus / Plus

```
Query:    590 GGTGCAGGCGGCTGACATTGTGAGCCGGGTCAACTGGCTGGGCCATCTCGGGCAGCCTCT 531
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    133 GGTGCAGGCGGCTGACATTGTGAGCCGGGTCAACTGGCTGGGCCATCTCGGGCAGCCTCT 192

Query:    530 TTCTTCGGGCAGGCTTGGCCACCTCATGGTCTAGGT 495
              ||||||||||||||||||||||||||||||||||||
Sbjct:    193 TTCTTCGGGCAGGCTTGGCCACCTCATGGTCTAGGT 228
```

Score = 410 (61.5 bits), Expect = 2.7e-31, Sum P(2) = 2.7e-31 **(SEQ ID NO:132)**

```
 Identities = 86/91 (94%), Positives = 86/91 (94%), Strand = Minus / Plus

Query:   681 GGATCCCCAAACCCCGCCTTGTAGGCTTCCCAGGGCCGGTTGAAGTCCACTGAGCCATCG 622
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:     1 GGATCCCCAAACCCCGCCTTGTAGGCTTCCCAGGGCCGGTTGAAGTCCACTGAGCCATCG 60

Query:   621 TGGCGCCTCTGAATTACTGTCCAGCCTCCAT 591
             |||||||||||||||| ||||||  || | |
Sbjct:    61 TGGCGCCTCTGAATTAATGTCCACTCTGCCT 91
```

Table 37. ClustalW alignment of CG57051-03 protein with related proteins.

```
CG57051-03  MSGAPTAGAALMLCAATAVLLSAQGGPVQSKSPRFASWDEMNVLAHGLLQLGQGLREHAE
Q9HBV4      MSGAPTAGAALMLCAATAVLLSAQGGPVQSKSPRFASWDEMNVLAHGLLQLGQGLREHAE
CG57051-02  MSGAPTAGAALMLCAATAVLLSARSGPVQSKSPRFASWDEMNVLAHGLLQLGQGLREHAE

CG57051-03  RTRSQLSALERRLSACGSACQGTEGSTDLPLAPESRVDPEVLHSLQTQLKAQNSRIQQLF
Q9HBV4      RTRSQLSALERRLSACGSACQGTEGSTDLPLAPESRVDPEVLHSLQTQLKAQNSRIQQLF
CG57051-02  RTRSQLSALERRLSACGSACQGTEGSTDLPLAPESRVDPEVLHSLQTQLKAQNSRIQQLF

CG57051-03  HKVAQQQRHLEKQHLRIQHLQSQFGLLDHKHLDHEVAKPARRKRLPEMAQPVDPAHNVSR
Q9HBV4      HKVAQQQRHLEKQHLRIQHLQSQFGLLDHKHLDHEVAKPARRKRLPEMAQPVDPAHNVSR
CG57051-02  HKVAQQQRHLEKQHLRIQHLQSQFGLLDHKHLDHEVAKPARRKRLPEMAQPVDPAHNVSR

CG57051-03  LH----------------------------------------HGGWTVIQRRHDGSVDFNRP
Q9HBV4      LHRLPRDCQELFQVGERQSGLFEIQPQGSPPFLVNCKMTSDGGWTVIQRRHDGSVDFNRP
CG57051-02  LH----------------------------------------HGGWTVIQRRHDGSMDFNRP

CG57051-03  WEAYKAGFGDPHGEFWLGLEKVHSITGDRNSRLAVQLRDWDDNAELLQFSVHLGGEDTAY
Q9HBV4      WEAYKAGFGDPHGEFWLGLEKVHSITGDRNSRLAVQLRDWDGNAELLQFSVHLGGEDTAY
CG57051-02  WEAYKAGFGDPHGEFWLGLEKVHSITGDRNSRLAVQLRDWDGNAELLQFSVHLGGEDTAY

CG57051-03  SLQLTAPVAGQLGATTVPPSGLSVPFPTWDQDHDLRRDKNCAKSLS--------------
Q9HBV4      SLQLTAPVAGQLGATTVPPSGLSVPFSTWDQDHDLRRDKNCAKSLS--------------
CG57051-02  SLQLTAPVAGQLGATTVPPSGLSVPFSTWDQDHDLRRDKNCAKSLSAPSVAQRPDHVPSP

CG57051-03  ----GGWWFGTCSHSNLNGQYFRSIPQQRQKLKKGIFWKTWRGRYYPLQATTMLIQPMAA
Q9HBV4      ----GGWWFGTCSHSNLNGQYFRSIPQQRQKLKKGIFWKTWRGRYYPLQATTMLIQPMAA
CG57051-02  LTPAGGWWFGTCSHSNLNGQYFRSIPQQRQKLKKGIFWKTWRGRYYPLQATTMLIQPMAA

CG57051-03  EAAS
Q9HBV4      EAAS
CG57051-02  EAAS
```

Information for the ClustalW proteins:

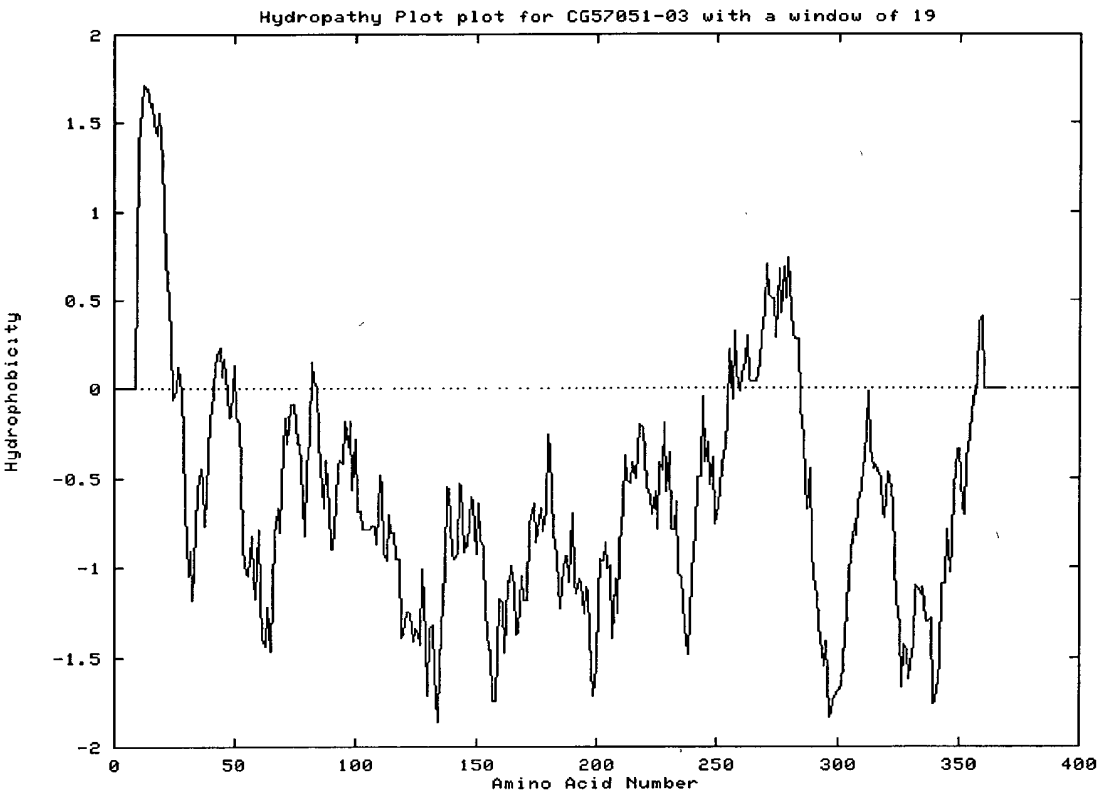
Accno	Common Name	Length
CG57051-03 (SEQ ID NO:49)	novel Angiopoietin-like protein	368
Q9HBV4 (SEQ ID NO:80)	ANGIOPOIETIN-LIKE PROTEIN PP1158.	406
CG57051-02 (SEQ ID NO:55)	Angiopoietin-like protein- isoform 2	386

In the alignment shown above, black outlined amino acid residues indicate residues identically conserved between sequences (i.e., residues that may be required to preserve

structural or functional properties); amino acid residues with a gray background are similar to one another between sequences, possessing comparable physical and/or chemical properties without altering protein structure or function (e.g. the group L,V, I, and M may be considered similar); and amino acid residues with a white background are neither conserved nor similar between sequences.

Table 38. PSORT, SignalP and hydropathy results for CuraGen Acc. No. CG57051-03.

```
                    outside --- Certainty=0.7332(Affirmative) < succ>
          microbody (peroxisome) --- Certainty=0.2527(Affirmative) < succ>
 endoplasmic reticulum (membrane) --- Certainty=0.1000(Affirmative) < succ>
    endoplasmic reticulum (lumen) --- Certainty=0.1000(Affirmative) < succ>


Is the sequence a signal peptide?
# Measure  Position  Value  Cutoff  Conclusion
  max. C    31        0.306   0.37   NO
  max. Y    26        0.429   0.34   YES
  max. S     8        0.952   0.88   YES
  mean S     1-25     0.848   0.48   YES
# Most likely cleavage site between pos. 25 and 26: AQG-GP
```



CG57051-04 directed toward novel Angiopoietin-like proteins and nucleic acids encoding them. Figure 20 illustrates the nucleic acid sequence and amino acid sequences respectively. This clone includes a nucleotide sequence (SEQ ID NO:50) of 937 bp. The nucleotide sequence includes an open reading frame (ORF) beginning with an ATG initiation codon at nucleotides 155-157 and ending with a TAG stop codon at nucleotides 881-883. Putative untranslated regions, if any, are found upstream from the initiation codon and downstream from the termination codon. The encoded protein having 242 amino acid residues is presented using the one-letter code in Figure 20. The protein encoded by clone CG57051-04 is predicted by the PSORT program to be located at the endoplasmic reticulum with a certainty of 0.8200, and appears to be a signal protein (see Table 27 below). Bottom of Form

SECP Nucleic Acids

The novel nucleic acids of the invention include those that encode a SECP or SECP-like protein, or biologically-active portions thereof. The nucleic acids include nucleic acids encoding polypeptides that include the amino acid sequence of one or more of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and 56. The encoded polypeptides can thus include, *e.g.*, the amino acid sequences of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and 56.In some embodiments, a SECP polypeptide or protein, as disclosed herein, includes the product of a naturally-occurring polypeptide, precursor form, pro-protein, or mature form of the polypeptide. The naturally-occurring polypeptide, precursor, or pro-protein includes, *e.g.*, the full-length gene product, encoded by the corresponding gene. The naturally-occurring polypeptide also includes the polypeptide, precursor or pro-protein encoded by an open reading frame (ORF) described herein. As used herein, the term "identical" residues corresponds to those residues in a comparison between two sequences where the equivalent nucleotide base or amino acid residue in an alignment of two sequences is the same residue. Residues are alternatively described as "similar" or "positive" when the comparisons between two sequences in an alignment show that residues in an equivalent position in a comparison are either the same amino acid residue or a conserved amino acid residue, as defined below.

As used herein, a "mature" form of a polypeptide or protein disclosed in the present invention is the product of a naturally occurring polypeptide or precursor form or proprotein. The naturally occurring polypeptide, precursor or proprotein includes, by way of nonlimiting example, the full length gene product, encoded by the corresponding gene. Alternatively, it may be defined as the polypeptide, precursor or proprotein encoded by an open reading frame

described herein. The product "mature" form arises, again by way of nonlimiting example, as a result of one or more naturally occurring processing steps as they may take place within the cell, or host cell, in which the gene product arises. Examples of such processing steps leading to a "mature" form of a polypeptide or protein include the cleavage of the amino-terminal methionine residue encoded by the initiation codon of an open reading frame, or the proteolytic cleavage of a signal peptide or leader sequence. Thus, a mature form arising from a precursor polypeptide or protein that has residues 1 to N, where residue 1 is the amino-terminal methionine, would have residues 2 through N remaining after removal of the amino-terminal methionine. Alternatively, a mature form arising from a precursor polypeptide or protein having residues 1 to N, in which an amino-terminal signal sequence from residue 1 to residue M is cleaved, would have the residues from residue M+1 to residue N remaining. Further, as used herein, a "mature" form of a polypeptide or protein may arise from a step of post-translational modification other than a proteolytic cleavage event. Such additional processes include, by way of non-limiting example, glycosylation, myristoylation or phosphorylation. In general, a mature polypeptide or protein may result from the operation of only one of these processes, or a combination of any of them.

In some embodiments, a nucleic acid encoding a polypeptide having the amino acid sequence of one or more of SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 41, 43, 45, 47, 49, 51, 53, 55 and 57, includes the nucleic acid sequence of any of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54, and 56, or a fragment thereof. Additionally, the invention includes mutant or variant nucleic acids of any of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and 56, or a fragment thereof, any of whose bases may be changed from the disclosed sequence while still encoding a protein that maintains its SECP-like biological activities and physiological functions. The invention further includes the complement of the nucleic acid sequence of any of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and 56, including fragments, derivatives, analogs and homologs thereof. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications.

Also included are nucleic acid fragments sufficient for use as hybridization probes to identify SECP-encoding nucleic acids (*e.g.*, SECP mRNA) and fragments for use as polymerase chain reaction (PCR) primers for the amplification or mutation of SECP nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA), RNA molecules (*e.g.*, mRNA), analogs of the DNA or RNA generated

using nucleotide analogs, and derivatives, fragments, and homologs thereof. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

The term "probes" refer to nucleic acid sequences of variable length, preferably between at least about 10 nucleotides (nt), 100 nt, or as many as about, *e.g.*, 6,000 nt, depending upon the specific use. Probes are used in the detection of identical, similar, or complementary nucleic acid sequences. Longer length probes are usually obtained from a natural or recombinant source, are highly specific and much slower to hybridize than oligomers. Probes may be single- or double-stranded, and may also be designed to have specificity in PCR, membrane-based hybridization technologies, or ELISA-like technologies.

The tem "isolated" nucleic acid molecule is a nucleic acid that is separated from other nucleic acid molecules that are present in the natural source of the nucleic acid. Examples of isolated nucleic acid molecules include, but are not limited to, recombinant DNA molecules contained in a vector, recombinant DNA molecules maintained in a heterologous host cell, partially or substantially purified nucleic acid molecules, and synthetic DNA or RNA molecules. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5'- and 3'-termini of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated SECP nucleic acid molecule can contain less than aprroximately 50 kb, 25 kb, 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material or culture medium when produced by recombinant techniques, or of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and 56, or a complement of any of these nucleotide sequences, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of any of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and 56 as a hybridization probe, SECP nucleic acid sequences can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook *et al.*, eds., MOLECULAR CLONING: A LABORATORY MANUAL 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989; and Ausubel, *et al.*, eds., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993.)

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to SECP nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

As used herein, the term "oligonucleotide" refers to a series of linked nucleotide residues, which oligonucleotide has a sufficient number of nucleotide bases to be used in a PCR reaction. A short oligonucleotide sequence may be based on, or designed from, a genomic or cDNA sequence and is used to amplify, confirm, or reveal the presence of an identical, similar or complementary DNA or RNA in a particular cell or tissue. Oligonucleotides comprise portions of a nucleic acid sequence having about 10 nt, 50 nt, or 100 nt in length, preferably about 15 nt to 30 nt in length. In one embodiment, an oligonucleotide comprising a nucleic acid molecule less than 100 nt in length would further comprise at lease 6 contiguous nucleotides of any of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and 56, or a complement thereof. Oligonucleotides may be chemically synthesized and may also be used as probes.

In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in any of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and 56. In still another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in any of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and 56, or a portion of this nucleotide sequence. A nucleic acid molecule that is complementary to the nucleotide sequence shown in is one that is sufficiently complementary to the nucleotide sequence shown in of any of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and 56 that it can hydrogen bond with little or no mismatches to the nucleotide sequence shown in of any of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and 56, thereby forming a stable duplex.

As used herein, the term "complementary" refers to Watson-Crick or Hoogsteen base-pairing between nucleotides units of a nucleic acid molecule, whereas the term "binding" is defined as the physical or chemical interaction between two polypeptides or compounds or associated polypeptides or compounds or combinations thereof. Binding includes ionic, non-ionic, Von der Waals, hydrophobic interactions, and the like. A physical interaction can be either direct or indirect. Indirect interactions may be through or due to the effects of another

polypeptide or compound. Direct binding refers to interactions that do not take place through, or due to, the effect of another polypeptide or compound, but instead are without other substantial chemical intermediates.

Additionally, the nucleic acid molecule of the invention can comprise only a portion of the nucleic acid sequence of any of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54, and 56,, *e.g.*, a fragment that can be used as a probe or primer, or a fragment encoding a biologically active portion of SECP. Fragments provided herein are defined as sequences of at least 6 (contiguous) nucleic acids or at least 4 (contiguous) amino acids, a length sufficient to allow for specific hybridization in the case of nucleic acids or for specific recognition of an epitope in the case of amino acids, respectively, and are at most some portion less than a full length sequence. Fragments may be derived from any contiguous portion of a nucleic acid or amino acid sequence of choice. Derivatives are nucleic acid sequences or amino acid sequences formed from the native compounds either directly or by modification or partial substitution. Analogs are nucleic acid sequences or amino acid sequences that have a structure similar to, but not identical to, the native compound but differs from it in respect to certain components or side chains. Analogs may be synthetic or from a different evolutionary origin and may have a similar or opposite metabolic activity compared to wild-type.

Derivatives and analogs may be full-length or other than full-length, if the derivative or analog contains a modified nucleic acid or amino acid, as described below. Derivatives or analogs of the nucleic acids or proteins of the invention include, but are not limited to, molecules comprising regions that are substantially homologous to the nucleic acids or proteins of the invention, in various embodiments, by at least about 70%, 80%, 85%, 90%, 95%, 98%, or even 99% identity (with a preferred identity of 80-99%) over a nucleic acid or amino acid sequence of identical size or when compared to an aligned sequence in which the alignment is done by a computer homology program known in the art, or whose encoding nucleic acid is capable of hybridizing to the complement of a sequence encoding the aforementioned proteins under stringent, moderately stringent, or low stringent conditions. See *e.g.* Ausubel, *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993, and below. An exemplary program is the Gap program (Wisconsin Sequence Analysis Package, Version 8 for UNIX, Genetics Computer Group, University Research Park, Madison, WI) using the default settings, which uses the algorithm of Smith and Waterman (Adv. Appl. Math., 1981, 2: 482-489), which is incorporated herein by reference in its entirety.

The tem "homologous nucleic acid sequence" or "homologous amino acid sequence," or variations thereof, refer to sequences characterized by a homology at the nucleotide level or amino acid level as previously discussed. Homologous nucleotide sequences encode those sequences coding for isoforms of SECP polypeptide. Isoforms can be expressed in different tissues of the same organism as a result of, *e.g.*, alternative splicing of RNA. Alternatively, isoforms can be encoded by different genes. In the invention, homologous nucleotide sequences include nucleotide sequences encoding for a SECP polypeptide of species other than humans, including, but not limited to, mammals, and thus can include, *e.g.*, mouse, rat, rabbit, dog, cat cow, horse, and other organisms. Homologous nucleotide sequences also include, but are not limited to, naturally occurring allelic variations and mutations of the nucleotide sequences set forth herein. A homologous nucleotide sequence does not, however, include the nucleotide sequence encoding human SECP protein. Homologous nucleic acid sequences include those nucleic acid sequences that encode conservative amino acid substitutions (see below) in any of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and 56, as well as a polypeptide having SECP activity. Biological activities of the SECP proteins are described below. A homologous amino acid sequence does not encode the amino acid sequence of a human SECP polypeptide.

The nucleotide sequence determined from the cloning of the human SECP gene allows for the generation of probes and primers designed for use in identifying the cell types disclosed and/or cloning SECP homologues in other cell types, *e.g.*, from other tissues, as well as SECP homologues from other mammals. The probe/primer typically comprises a substantially-purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 25, 50, 100, 150, 200, 250, 300, 350 or 400 or more consecutive sense strand nucleotide sequence of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and 56; or an anti-sense strand nucleotide sequence of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and 56, or of a naturally occurring mutant of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and 56.

Probes based upon the human SECP nucleotide sequence can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In various embodiments, the probe further comprises a label group attached thereto, *e.g.*, the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissue which mis-express a SECP

protein, such as by measuring a level of a SECP-encoding nucleic acid in a sample of cells from a subject *e.g.*, detecting SECP mRNA levels or determining whether a genomic SECP gene has been mutated or deleted.

The term "a polypeptide having a biologically-active portion of SECP" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. A nucleic acid fragment encoding a "biologically-active portion of SECP" can be prepared by isolating a portion of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and 56 that encodes a polypeptide having a SECP biological activity, expressing the encoded portion of SECP protein (*e.g.*, by recombinant expression *in vitro*), and assessing the activity of the encoded portion of SECP.

SECP Variants

The invention further encompasses nucleic acid molecules that differ from the disclosed SECP nucleotide sequences due to degeneracy of the genetic code. These nucleic acids therefore encode the same SECP protein as those encoded by the nucleotide sequence shown in SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and 56. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in any of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and 56.

In addition to the human SECP nucleotide sequence shown in any of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and 56, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of SECP may exist within a population (*e.g.*, the human population). Such genetic polymorphism in the SECP gene may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding a SECP protein, preferably a mammalian SECP protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the SECP gene. Any and all such nucleotide variations and resulting amino acid polymorphisms in SECP that are the result of natural allelic variation and that do not alter the functional activity of SECP are intended to be within the scope of the invention.

Additionally, nucleic acid molecules encoding SECP proteins from other species, and thus that have a nucleotide sequence that differs from the human sequence of any of SEQ ID

NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and 56 are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of the SECP cDNAs of the invention can be isolated based on their homology to the human SECP nucleic acids disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

In another embodiment, an isolated nucleic acid molecule of the invention is at least 6 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of any of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and/or 56. In another embodiment, the nucleic acid is at least 10, 25, 50, 100, 250, 500 or 750 nucleotides in length. In yet another embodiment, an isolated nucleic acid molecule of the invention hybridizes to the coding region. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other.

Homologs (*i.e.,* nucleic acids encoding SECP proteins derived from species other than human) or other related sequences (*e.g.,* paralogs) can be obtained by low, moderate or high stringency hybridization with all or a portion of the particular human sequence as a probe using methods well known in the art for nucleic acid hybridization and cloning.

As used herein, the phrase "stringent hybridization conditions" refers to conditions under which a probe, primer or oligonucleotide will hybridize to its target sequence, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures than shorter sequences. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. Since the target sequences are generally present at excess, at T_m, 50% of the probes are occupied at equilibrium. Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes, primers or oligonucleotides (*e.g.,* 10 nt to 50 nt) and at least about 60°C for longer probes, primers and oligonucleotides.

Stringent conditions may also be achieved with the addition of destabilizing agents, such as formamide.

Stringent conditions are known to those skilled in the art and can be found in CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Preferably, the conditions are such that sequences at least about 65%, 70%, 75%, 85%, 90%, 95%, 98%, or 99% homologous to each other typically remain hybridized to each other. A non-limiting example of stringent hybridization conditions is hybridization in a high salt buffer comprising 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 mg/ml denatured salmon sperm DNA at 65°C. This hybridization is followed by one or more washes in 0.2X SSC, 0.01% BSA at 50°C. An isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of any of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and 56 corresponds to a naturally occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein).

In a second embodiment, a nucleic acid sequence that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of any of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and/or 56, or fragments, analogs or derivatives thereof, under conditions of moderate stringency is provided. A non-limiting example of moderate stringency hybridization conditions are hybridization in 6X SSC, 5X Denhardt's solution, 0.5% SDS and 100 mg/ml denatured salmon sperm DNA at 55°C, followed by one or more washes in 1X SSC, 0.1% SDS at 37°C. Other conditions of moderate stringency that may be used are well known in the art. See, *e.g.*, Ausubel *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990. GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY.

In a third embodiment, a nucleic acid that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of any of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and 56, or fragments, analogs or derivatives thereof, under conditions of low stringency, is provided. A non-limiting example of low stringency hybridization conditions are hybridization in 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 mg/ml denatured salmon sperm DNA, 10% (wt/vol) dextran sulfate at 40°C, followed by one or more washes in 2X SSC, 25 mM Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS at 50°C. Other conditions of low stringency that may be used are well

known in the art (*e.g.*, as employed for cross-species hybridizations). *See, e.g.*, Ausubel, *et al.*, (eds.), 1993. CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990. GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY; Shilo and Weinberg, 1981. *Proc. Natl. Acad. Sci. USA 78*: 6789-6792.

Conservative Mutations

In addition to naturally-occurring allelic variants of the SECP sequence that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequence of any of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and 56, thereby leading to changes in the amino acid sequence of the encoded SECP protein, without altering the functional ability of the SECP protein. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence of any of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and 56. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of SECP without altering the biological activity, whereas an "essential" amino acid residue is required for biological activity. For example, amino acid residues that are conserved among the SECP proteins of the invention, are predicted to be particularly non-amenable to such alteration.

Amino acid residues that are conserved among members of a SECP family members are predicted to be less amenable to alteration. For example, a SECP protein according to the invention can contain at least one domain that is a typically conserved region in a SECP family member. As such, these conserved domains are not likely to be amenable to mutation. Other amino acid residues, however, (*e.g.*, those that are not conserved or only semi-conserved among members of the SECP family) may not be as essential for activity and thus are more likely to be amenable to alteration.

Another aspect of the invention pertains to nucleic acid molecules encoding SECP proteins that contain changes in amino acid residues that are not essential for activity. Such SECP proteins differ in amino acid sequence from any of any of SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 41, 43, 45, 47, 49, 51, 53, 55 and 57, yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 75% homologous to the amino acid sequence of any of SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 41, 43, 45, 47, 49, 51, 53, 55 and 57. Preferably, the protein encoded by the nucleic acid is at least about 80%

homologous to any of SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 41, 43, 45, 47, 49, 51, 53, 55 and 57, more preferably at least about 90%, 95%, 98%, and most preferably at least about 99% homologous to SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 41, 43, 45, 47, 49, 51, 53, 55 and 57,.

An isolated nucleic acid molecule encoding a SECP protein homologous to the protein of any of SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 41, 43, 45, 47, 49, 51, 53, 55 and 57, can be created by introducing one or more nucleotide substitutions, additions or deletions into the corresponding nucleotide sequence (*i.e.,* SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and/or 56), such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein.

Mutations can be introduced into SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and/or 56 by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), β-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in SECP is replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a SECP coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for SECP biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and/or 56, the encoded protein can be expressed by any recombinant technology known in the art and the activity of the protein can be determined.

In one embodiment, a mutant SECP protein can be assayed for: (*i*) the ability to form protein:protein interactions with other SECP proteins, other cell-surface proteins, or biologically-active portions thereof; (*ii*) complex formation between a mutant SECP protein and a SECP receptor; (*iii*) the ability of a mutant SECP protein to bind to an intracellular target protein or

biologically active portion thereof; (*e.g.*, avidin proteins); (*iv*) the ability to bind BRA protein; or (*v*) the ability to specifically bind an anti-SECP protein antibody.

Antisense Nucleic Acids

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and/or 56 or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire SECP coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of a SECP protein of any of SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 41, 43, 45, 47, 49, 51, 53, 55, and 57.

or antisense nucleic acids complementary to a SECP nucleic acid sequence of SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 41, 43, 45, 47, 49, 51, 53, 55 and 57, are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding SECP. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues (*e.g.*, the protein coding region of a human SECP that corresponds to any of SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 41, 43, 45, 47, 49, 51, 53, 55 and 57.

. In another embodiment, the antisense nucleic acid molecule is antisense to a "non-coding region" of the coding strand of a nucleotide sequence encoding SECP. The term "non-coding region" refers to 5'- and 3'-terminal sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' non-translated regions).

Given the coding strand sequences encoding the SECP proteins disclosed herein (*e.g.*, SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and/or 56), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base-pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of SECP mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or non-coding region of SECP mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of

SECP mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally-occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine-substituted nucleotides can be used.

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a SECP protein to thereby inhibit expression of the protein, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For

example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface (*e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens). The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual α-units, the strands run parallel to each other (*see,* Gaultier, *et al.*, 1987. *Nucl. Acids Res.* 15: 6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue, *et al.*, 1987. *Nucl. Acids Res.* 15: 6131-6148) or a chimeric RNA-DNA analogue (Inoue, *et al.*, 1987. *FEBS Lett.* 215: 327-330).

Ribozymes and PNA Moieties

Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes; described by Haselhoff and Gerlach, 1988. *Nature* 334: 585-591) can be used to catalytically-cleave SECP mRNA transcripts to thereby inhibit translation of SECP mRNA. A ribozyme having specificity for a SECP-encoding nucleic acid can be designed based upon the nucleotide sequence of a SECP DNA disclosed herein (*i.e.*, SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and/or 56). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a SECP-encoding mRNA. See, *e.g.*, Cech, *et al.*, U.S. Patent No. 4,987,071; and Cech, *et al.*, U.S. Patent No. 5,116,742. Alternatively, SECP mRNA can be used to select a catalytic RNA having a specific

ribonuclease activity from a pool of RNA molecules (Bartel, *et al.*, 1993. *Science* 261: 1411-1418).

Alternatively, SECP gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the SECP (*e.g.*, the SECP promoter and/or enhancers) to form triple helical structures that prevent transcription of the SECP gene in target cells. *See, e.g.,* Helene, 1991. *Anticancer Drug Des.* 6: 569-84; Helene, *et al.*, 1992. *Ann. N.Y. Acad. Sci.* 660: 27-36; and Maher, 1992. *Bioassays* 14: 807-15.

In various embodiments, the nucleic acids of SECP can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (Hyrup, *et al.*, 1996. *Bioorg. Med. Chem.* 4: 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup, *et al.*, 1996. *supra*; Perry-O'Keefe, *et al.*, 1996. *Proc. Natl. Acad. Sci. USA* 93: 14670-14675.

PNAs of SECP can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs of SECP can also be used, *e.g.*, in the analysis of single base pair mutations in a gene by, *e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S1 nucleases (*see,* Hyrup, 1996., *supra*); or as probes or primers for DNA sequence and hybridization (*see,* Hyrup, *et al.*, 1996.; Perry-O'Keefe, 1996., *supra*).

In another embodiment, PNAs of SECP can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of SECP can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, *e.g.*, RNase H and DNA polymerases, to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using

linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (*see,* Hyrup, 1996., *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Finn, *et al.*, (1996. *Nucl. Acids Res.* 24: 3357-3363). For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA (Mag, *et al.*, 1989. *Nucl. Acid Res.* 17: 5973-5988). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (*see,* Finn, *et al.*, 1996., *supra*). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. *See. e.g.*. Petersen, *et al.*, 1975. *Bioorg. Med. Chem. Lett.* 5: 1119-11124.

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (*see, e.g.*, Letsinger, *et al.*, 1989. *Proc. Natl. Acad. Sci. U.S.A.* 86: 6553-6556; Lemaitre, *et al.*, 1987. *Proc. Natl. Acad. Sci.* 84: 648-652; PCT Publication No. WO88/09810) or the blood-brain barrier (*see, e.g.*, PCT Publication No. WO 89/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents (*see, e.g.*, Krol, *et al.*, 1988. *BioTechniques* 6:958-976) or intercalating agents (*see, e.g.*, Zon, 1988. *Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, and the like.

Characterization of SECP Polypeptides

A polypeptide according to the invention includes a polypeptide including the amino acid sequence of SECP polypeptides whose sequences are provided in SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 41, 43, 45, 47, 49, 51, 53, 55, and/or 57. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residues shown in SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 41, 43, 45, 47, 49, 51, 53, 55, and/or 57 while still encoding a protein that maintains its SECP activities and physiological functions, or a functional fragment thereof.

In general, a SECP variant that preserves SECP-like function includes any variant in which residues at a particular position in the sequence have been substituted by other amino acids, and further include the possibility of inserting an additional residue or residues between

two residues of the parent protein as well as the possibility of deleting one or more residues from the parent sequence. Any amino acid substitution, insertion, or deletion is encompassed by the invention. In favorable circumstances, the substitution is a conservative substitution as defined above.

One aspect of the invention pertains to isolated SECP proteins, and biologically-active portions thereof, or derivatives, fragments, analogs or homologs thereof. Also provided are polypeptide fragments suitable for use as immunogens to raise anti-SECP antibodies. In one embodiment, native SECP proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, SECP proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, a SECP protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" polypeptide or protein or biologically-active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the SECP protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of SECP proteins in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly-produced. In one embodiment, the language "substantially free of cellular material" includes preparations of SECP proteins having less than about 30% (by dry weight) of non-SECP proteins (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-SECP proteins, still more preferably less than about 10% of non-SECP proteins, and most preferably less than about 5% of non-SECP proteins. When the SECP protein or biologically-active portion thereof is recombinantly-produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the SECP protein preparation.

The phrase "substantially free of chemical precursors or other chemicals" includes preparations of SECP protein in which the protein is separated from chemical precursors or other chemicals that are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of SECP protein having less than about 30% (by dry weight) of chemical precursors or non-SECP chemicals, more preferably less than about 20% chemical precursors or non-SECP chemicals,

still more preferably less than about 10% chemical precursors or non-SECP chemicals, and most preferably less than about 5% chemical precursors or non-SECP chemicals.

Biologically-active portions of a SECP protein include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequence of the SECP protein which include fewer amino acids than the full-length SECP proteins, and exhibit at least one activity of a SECP protein. Typically, biologically-active portions comprise a domain or motif with at least one activity of the SECP protein. A biologically-active portion of a SECP protein can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acids in length.

A biologically-active portion of a SECP protein of the invention may contain at least one of the above-identified conserved domains. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native SECP protein.

In an embodiment, the SECP protein has an amino acid sequence shown in any of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and/or 56. In other embodiments, the SECP protein is substantially homologous to any of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and/or 56 and retains the functional activity of the protein of any of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and/or 56 yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail below. Accordingly, in another embodiment, the SECP protein is a protein that comprises an amino acid sequence at least about 45% homologous, and more preferably about 55, 65, 70, 75, 80, 85, 90, 95, 98 or even 99% homologous to the amino acid sequence of any of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and/or 56 and retains the functional activity of the SECP proteins of the corresponding polypeptide having the sequence of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and/or 56.

Determining Homology Between Two or More Sequences

To determine the percent homology of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the

second sequence, then the molecules are homologous at that position (*i.e.*, as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity").

The nucleic acid sequence homology may be determined as the degree of identity between two sequences. The homology may be determined using computer programs known in the art, such as GAP software provided in the GCG program package. *See,* Needleman and Wunsch, 1970. *J. Mol. Biol.* 48: 443-453. Using GCG GAP software with the following settings for nucleic acid sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3, the coding region of the analogous nucleic acid sequences referred to above exhibits a degree of identity preferably of at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99%, with the CDS (encoding) part of the DNA sequence shown in SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and/or 56.

The term "sequence identity" refers to the degree to which two polynucleotide or polypeptide sequences are identical on a residue-by-residue basis over a particular region of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of positions at which the identical nucleic acid base (*e.g.*, A, T, C, G, U, or I, in the case of nucleic acids) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 80 percent sequence identity, preferably at least 85 percent identity and often 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison region.

Chimeric and Fusion Proteins

The invention also provides SECP chimeric or fusion proteins. As used herein, a SECP "chimeric protein" or "fusion protein" comprises a SECP polypeptide operatively-linked to a non-SECP polypeptide. An "SECP polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a SECP protein shown in SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 41, 43, 45, 47, 49, 51, 53, 55, and/or 57, whereas a "non-SECP polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein that is not substantially homologous to the SECP protein (*e.g.*, a protein that is different from the SECP protein and that

is derived from the same or a different organism). Within a SECP fusion protein the SECP polypeptide can correspond to all or a portion of a SECP protein. In one embodiment, a SECP fusion protein comprises at least one biologically-active portion of a SECP protein. In another embodiment, a SECP fusion protein comprises at least two biologically-active portions of a SECP protein. In yet another embodiment, a SECP fusion protein comprises at least three biologically-active portions of a SECP protein. Within the fusion protein, the term "operatively-linked" is intended to indicate that the SECP polypeptide and the non-SECP polypeptide are fused in-frame with one another. The non-SECP polypeptide can be fused to the amino-terminus or carboxyl-terminus of the SECP polypeptide.

In one embodiment, the fusion protein is a GST-SECP fusion protein in which the SECP sequences are fused to the carboxyl-terminus of the GST (glutathione S-transferase) sequences. Such fusion proteins can facilitate the purification of recombinant SECP polypeptides.

In another embodiment, the fusion protein is a SECP protein containing a heterologous signal sequence at its amino-terminus. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of SECP can be increased through use of a heterologous signal sequence.

In yet another embodiment, the fusion protein is a SECP-immunoglobulin fusion protein in which the SECP sequences are fused to sequences derived from a member of the immunoglobulin protein family. The SECP-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a SECP ligand and a SECP protein on the surface of a cell, to thereby suppress SECP-mediated signal transduction *in vivo*. The SECP-immunoglobulin fusion proteins can be used to affect the bioavailability of a SECP cognate ligand. Inhibition of the SECP ligand/SECP interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, as well as modulating (*e.g.*, promoting or inhibiting) cell survival. Moreover, the SECP-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-SECP antibodies in a subject, to purify SECP ligands, and in screening assays to identify molecules that inhibit the interaction of SECP with a SECP ligand.

A SECP chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme

digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (*see, e.g.,* Ausubel, *et al.* (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). A SECP-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the SECP protein.

SECP Agonists and Antagonists

The invention also pertains to variants of the SECP proteins that function as either SECP agonists (*i.e.,* mimetics) or as SECP antagonists. Variants of the SECP protein can be generated by mutagenesis (*e.g.*, discrete point mutation or truncation of the SECP protein). An agonist of a SECP protein can retain substantially the same, or a subset of, the biological activities of the naturally-occurring form of a SECP protein. An antagonist of a SECP protein can inhibit one or more of the activities of the naturally occurring form of a SECP protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the SECP protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the SECP proteins.

Variants of the SECP proteins that function as either SECP agonists (*i.e.,* mimetics) or as SECP antagonists can be identified by screening combinatorial libraries of mutants (*e.g.*, truncation mutants) of the SECP proteins for SECP protein agonist or antagonist activity. In one embodiment, a variegated library of SECP variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of SECP variants can be produced by, for example, enzymatically-ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential SECP sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display) containing the set of SECP sequences therein. There are a variety of methods

which can be used to produce libraries of potential SECP variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential SECP sequences. Methods for synthesizing degenerate oligonucleotides are well-known within the art. *See, e.g.,* Narang, 1983. *Tetrahedron* 39: 3; Itakura, *et al.*, 1984. *Annu. Rev. Biochem.* 53: 323; Itakura, *et al.*, 1984. *Science* 198: 1056; Ike, *et al.*, 1983. *Nucl. Acids Res.* 11: 477.

Polypeptide Libraries

In addition, libraries of fragments of the SECP protein coding sequences can be used to generate a variegated population of SECP fragments for screening and subsequent selection of variants of a SECP protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double-stranded PCR fragment of a SECP coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double-stranded DNA that can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S_1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, expression libraries can be derived which encodes amino-terminal and internal fragments of various sizes of the SECP proteins.

Various techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of SECP proteins. The most widely used techniques, which are amenable to high throughput analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique that enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify SECP variants. *See, e.g.,* Arkin and Yourvan, 1992. *Proc. Natl. Acad. Sci. USA* 89: 7811-7815; Delgrave, *et al.*, 1993. *Protein Engineering* 6:327-331.

Anti-SECP Antibodies

The invention encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the SECP polypeptides of said invention.

An isolated SECP protein, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that bind to SECP polypeptides using standard techniques for polyclonal and monoclonal antibody preparation. The full-length SECP proteins can be used or, alternatively, the invention provides antigenic peptide fragments of SECP proteins for use as immunogens. The antigenic SECP peptides comprises at least 4 amino acid residues of the amino acid sequence shown in SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 41, 43, 45, 47, 49, 51, 53, 55, and/or 57, and encompasses an epitope of SECP such that an antibody raised against the peptide forms a specific immune complex with SECP. Preferably, the antigenic peptide comprises at least 6, 8, 10, 15, 20, or 30 amino acid residues. Longer antigenic peptides are sometimes preferable over shorter antigenic peptides, depending on use and according to methods well known to someone skilled in the art.

In certain embodiments of the invention, at least one epitope encompassed by the antigenic peptide is a region of SECP that is located on the surface of the protein (*e.g.*, a hydrophilic region). As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and hydrophobicity may be generated by any method well known in the art, including, for example, the Kyte-Doolittle or the Hopp-Woods methods, either with or without Fourier transformation (*see, e.g.*, Hopp and Woods, 1981. *Proc. Nat. Acad. Sci. USA* 78: 3824-3828; Kyte and Doolittle, 1982. *J. Mol. Biol.* 157: 105-142, each incorporated herein by reference in their entirety).

As disclosed herein, SECP protein sequences of SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 41, 43, 45, 47, 49, 51, 53, 55, and/or 57, or derivatives, fragments, analogs, or homologs thereof, may be utilized as immunogens in the generation of antibodies that immunospecifically-bind these protein components. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically-active portions of immunoglobulin molecules, *i.e.*, molecules that contain an antigen binding site that specifically-binds (immunoreacts with) an antigen, such as SECP. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, F_{ab} and $F_{(ab')2}$ fragments, and an F_{ab} expression library. In a specific embodiment, antibodies to human SECP proteins are disclosed. Various procedures known within the art may be used for the production of polyclonal or monoclonal antibodies to a SECP protein sequence of

SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 41, 43, 45, 47, 49, 51, 53, 55 and/or 57, or a derivative, fragment, analog, or homolog thereof.

For the production of polyclonal antibodies, various suitable host animals (*e.g.*, rabbit, goat, mouse or other mammal) may be immunized by injection with the native protein, or a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, recombinantly-expressed SECP protein or a chemically-synthesized SECP polypeptide. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete and incomplete), mineral gels (*e.g.*, aluminum hydroxide), surface active substances (*e.g.*, lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, etc.), human adjuvants such as *Bacille Calmette-Guerin* and *Corynebacterium parvum*, or similar immunostimulatory agents. If desired, the antibody molecules directed against SECP can be isolated from the mammal (*e.g.*, from the blood) and further purified by well known techniques, such as protein A chromatography to obtain the IgG fraction.

The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of SECP. A monoclonal antibody composition thus typically displays a single binding affinity for a particular SECP protein with which it immunoreacts. For preparation of monoclonal antibodies directed towards a particular SECP protein, or derivatives, fragments, analogs or homologs thereof, any technique that provides for the production of antibody molecules by continuous cell line culture may be utilized. Such techniques include, but are not limited to, the hybridoma technique (*see, e.g.,* Kohler & Milstein, 1975. *Nature* 256: 495-497); the trioma technique; the human B-cell hybridoma technique (*see, e.g.,* Kozbor, *et al.,* 1983. *Immunol. Today* 4: 72) and the EBV hybridoma technique to produce human monoclonal antibodies (*see, e.g.,* Cole, *et al.,* 1985. In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the invention and may be produced by using human hybridomas (*see, e.g.,* Cote, *et al.,* 1983. *Proc Natl Acad Sci USA* 80: 2026-2030) or by transforming human B-cells with Epstein Barr Virus *in vitro* (*see, e.g.,* Cole, *et al.,* 1985. In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Each of the above citations is incorporated herein by reference in their entirety.

According to the invention, techniques can be adapted for the production of single-chain antibodies specific to a SECP protein (*see, e.g.,* U.S. Patent No. 4,946,778). In addition,

methods can be adapted for the construction of F_{ab} expression libraries (*see, e.g.,* Huse, *et al.,* 1989. *Science* 246: 1275-1281) to allow rapid and effective identification of monoclonal F_{ab} fragments with the desired specificity for a SECP protein or derivatives, fragments, analogs or homologs thereof. Non-human antibodies can be "humanized" by techniques well known in the art. *See, e.g.,* U.S. Patent No. 5,225,539. Antibody fragments that contain the idiotypes to a SECP protein may be produced by techniques known in the art including, but not limited to: (*i*) an $F_{(ab')2}$ fragment produced by pepsin digestion of an antibody molecule; (*ii*) an F_{ab} fragment generated by reducing the disulfide bridges of an $F_{(ab')2}$ fragment; (*iii*) an F_{ab} fragment generated by the treatment of the antibody molecule with papain and a reducing agent and (*iv*) F_v fragments.

Additionally, recombinant anti-SECP antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in International Application No. PCT/US86/02269; European Patent Application No. 184,187; European Patent Application No. 171,496; European Patent Application No. 173,494; PCT International Publication No. WO 86/01533; U.S. Patent No. 4,816,567; U.S. Pat. No. 5,225,539; European Patent Application No. 125,023; Better, *et al.,* 1988. *Science* 240: 1041-1043; Liu, *et al.,* 1987. *Proc. Natl. Acad. Sci. USA* 84: 3439-3443; Liu, *et al.,* 1987. *J. Immunol.* 139: 3521-3526; Sun, *et al.,* 1987. *Proc. Natl. Acad. Sci. USA* 84: 214-218; Nishimura, *et al.,* 1987. *Cancer Res.* 47: 999-1005; Wood, *et al.,* 1985. *Nature* 314 :446-449; Shaw, *et al.,* 1988. *J. Natl. Cancer Inst.* 80: 1553-1559); Morrison(1985) *Science* 229:1202-1207; Oi, *et al.* (1986) *BioTechniques* 4:214; Jones, *et al.,* 1986. *Nature* 321: 552-525; Verhoeyan, *et al.,* 1988. *Science* 239: 1534; and Beidler, *et al.,* 1988. *J. Immunol.* 141: 4053-4060. Each of the above citations are incorporated herein by reference in their entirety.

In one embodiment, methods for the screening of antibodies that possess the desired specificity include, but are not limited to, enzyme-linked immunosorbent assay (ELISA) and other immunologically-mediated techniques known within the art. In a specific embodiment, selection of antibodies that are specific to a particular domain of a SECP protein is facilitated by generation of hybridomas that bind to the fragment of a SECP protein possessing such a domain. Thus, antibodies that are specific for a desired domain within a SECP protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

Anti-SECP antibodies may be used in methods known within the art relating to the localization and/or quantitation of a SECP protein (*e.g.,* for use in measuring levels of the SECP protein within appropriate physiological samples, for use in diagnostic methods, for use in imaging the protein, and the like). In a given embodiment, antibodies for SECP proteins, or derivatives, fragments, analogs or homologs thereof, that contain the antibody derived binding domain, are utilized as pharmacologically-active compounds (hereinafter "Therapeutics").

An anti-SECP antibody (*e.g.*, monoclonal antibody) can be used to isolate a SECP polypeptide by standard techniques, such as affinity chromatography or immunoprecipitation. An anti-SECP antibody can facilitate the purification of natural SECP polypeptide from cells and of recombinantly-produced SECP polypeptide expressed in host cells. Moreover, an anti-SECP antibody can be used to detect SECP protein (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the SECP protein. Anti-SECP antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (*i.e.*, physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

SECP Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding a SECP protein, or derivatives, fragments, analogs or homologs thereof. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host

cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present Specification, "plasmid" and "vector" can be used interchangeably, as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably-linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner that allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell).

The phrase "regulatory sequence" is intended to includes promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (*e.g.*, SECP proteins, mutant forms of SECP proteins, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of SECP proteins in prokaryotic or eukaryotic cells. For example, SECP proteins can be expressed

in bacterial cells such as *Escherichia coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T_7 promoter regulatory sequences and T_7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *Escherichia coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: (*i*) to increase expression of recombinant protein; (*ii*) to increase the solubility of the recombinant protein; and (*iii*) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor X_a, thrombin, and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988. *Gene* 67: 31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion *Escherichia coli* expression vectors include pTrc (Amrann *et al.*, (1988) *Gene* 69:301-315) and pET 11d (Studier, *et al.*, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89).

One strategy to maximize recombinant protein expression in *Escherichia coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically-cleave the recombinant protein. *See, e.g.,* Gottesman, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 119-128. Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *Escherichia coli* (*see, e.g.,* Wada, *et al.*, 1992. *Nucl. Acids Res.* 20: 2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the SECP expression vector is a yeast expression vector. Examples of vectors for expression in yeast *Saccharomyces cerivisae* include pYepSec1 (Baldari, *et al.*, 1987. *EMBO J.* 6: 229-234), pMFa (Kurjan and Herskowitz, 1982. *Cell* 30: 933-943), pJRY88 (Schultz *et al.*, 1987. *Gene* 54: 113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (InVitrogen, Corp.; San Diego, Calif.).

Alternatively, SECP can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*, SF9 cells) include the pAc series (Smith, *et al.*, 1983. *Mol. Cell. Biol.* 3: 2156-2165) and the pVL series (Lucklow and Summers, 1989. *Virology* 170: 31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, 1987. *Nature* 329: 840) and pMT2PC (Kaufman, *et al.*, 1987. *EMBO J.* 6: 187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, adenovirus 2, cytomegalovirus, and simian virus 40 (SV 40). For other suitable expression systems for both prokaryotic and eukaryotic cells *see, e.g.*, Chapters 16 and 17 of Sambrook, *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; *see,* Pinkert, *et al.*, 1987. *Genes Dev.* 1: 268-277), lymphoid-specific promoters (*see,* Calame and Eaton, 1988. *Adv. Immunol.* 43: 235-275), in particular promoters of T cell receptors (*see,* Winoto and Baltimore, 1989. *EMBO J.* 8: 729-733) and immunoglobulins (*see,* Banerji, *et al.*, 1983. *Cell* 33: 729-740; Queen and Baltimore, 1983. *Cell* 33: 741-748), neuron-specific promoters (*e.g.*, the neurofilament promoter; *see,* Byrne and Ruddle, 1989. *Proc. Natl. Acad. Sci. USA* 86: 5473-5477), pancreas-specific promoters (*see,* Edlund, *et al.*, 1985. *Science* 230: 912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, *e.g.*, the

murine hox promoters (Kessel and Gruss, 1990. *Science* 249: 374-379) and the α-fetoprotein promoter (*see,* Campes and Tilghman, 1989. *Genes Dev.* 3: 537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively-linked to a regulatory sequence in a manner that allows for expression (by transcription of the DNA molecule) of an RNA molecule that is antisense to SECP mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen that direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen that direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes *see, e.g.,* Weintraub, *et al.*, "Antisense RNA as a molecular tool for genetic analysis," *Reviews-Trends in Genetics*, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, SECP protein can be expressed in bacterial cells such as *Escherichia coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation.

Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Various selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding SECP or can be introduced on a separate vector. Cells stably-transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) SECP protein. Accordingly, the invention further provides methods for producing SECP protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (*i.e.*, into which a recombinant expression vector encoding SECP protein has been introduced) in a suitable medium such that SECP protein is produced. In another embodiment, the method further comprises isolating SECP protein from the medium or the host cell.

Transgenic Animals

The host cells of the invention can also be used to produce non-human transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which SECP protein-coding sequences have been introduced. These host cells can then be used to create non-human transgenic animals in which exogenous SECP sequences have been introduced into their genome or homologous recombinant animals in which endogenous SECP sequences have been altered. Such animals are useful for studying the function and/or activity of SECP protein and for identifying and/or evaluating modulators of SECP protein activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of

the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, etc.

A transgene is exogenous DNA that is integrated into the genome of a cell from which a transgenic animal develops and that remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous SECP gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing SECP-encoding nucleic acid into the male pronuclei of a fertilized oocyte (*e.g.*, by micro-injection, retroviral infection) and allowing the oocyte to develop in a pseudopregnant female foster animal. The human SECP cDNA sequences of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and/or 56 can be introduced as a transgene into the genome of a non-human animal. Alternatively, a non-human homologue of the human SECP gene, such as a mouse SECP gene, can be isolated based on hybridization to the human SECP cDNA (described further *supra*) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably-linked to the SECP transgene to direct expression of SECP protein to particular cells. Methods for generating transgenic animals via embryo manipulation and micro-injection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866; 4,870,009; and 4,873,191; and Hogan, 1986. In: MANIPULATING THE MOUSE EMBRYO, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the SECP transgene in its genome and/or expression of SECP mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene-encoding SECP protein can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of a SECP gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the SECP gene. The SECP gene can be a human gene

(*e.g.*, the cDNA of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and 56), but more preferably, is a non-human homologue of a human SECP gene. For example, a mouse homologue of human SECP gene of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and 56 can be used to construct a homologous recombination vector suitable for altering an endogenous SECP gene in the mouse genome. In one embodiment, the vector is designed such that, upon homologous recombination, the endogenous SECP gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector).

Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous SECP gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the endogenous SECP protein). In the homologous recombination vector, the altered portion of the SECP gene is flanked at its 5'- and 3'-termini by additional nucleic acid of the SECP gene to allow for homologous recombination to occur between the exogenous SECP gene carried by the vector and an endogenous SECP gene in an embryonic stem cell. The additional flanking SECP nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases (Kb) of flanking DNA (both at the 5'- and 3'-termini) are included in the vector. *See, e.g.,* Thomas, *et al.*, 1987. *Cell* 51: 503 for a description of homologous recombination vectors. The vector is ten introduced into an embryonic stem cell line (*e.g.*, by electroporation) and cells in which the introduced SECP gene has homologously-recombined with the endogenous SECP gene are selected. *See, e.g.*, Li, *et al.*, 1992. *Cell* 69: 915.

The selected cells are then micro-injected into a blastocyst of an animal (*e.g.*, a mouse) to form aggregation chimeras. *See, e.g.*, Bradley, 1987. In: TERATOCARCINOMAS AND EMBRYONIC STEM CELLS: A PRACTICAL APPROACH, Robertson, ed. IRL, Oxford, pp. 113-152. A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously-recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously-recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley, 1991. *Curr. Opin. Biotechnol.* 2: 823-829; PCT International Publication Nos.: WO 90/11354; WO 91/01140; WO 92/0968; and WO 93/04169.

In another embodiment, transgenic non-human animals can be produced that contain selected systems that allow for regulated expression of the transgene. One example of such a system is the cre/loxP recombinase system of bacteriophage P1. For a description of the cre/loxP

recombinase system, *See, e.g.*, Lakso, *et al.*, 1992. *Proc. Natl. Acad. Sci. USA* 89: 6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae. See,* O'Gorman, *et al.*, 1991. *Science* 251:1351-1355. If a cre/loxP recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the Cre recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, *e.g.*, by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, *et al.*, 1997. *Nature* 385: 810-813. In brief, a cell (*e.g.*, a somatic cell) from the transgenic animal can be isolated and induced to exit the growth cycle and enter G_0 phase. The quiescent cell can then be fused, *e.g.*, through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell (*e.g.*, the somatic cell) is isolated.

Pharmaceutical Compositions

The SECP nucleic acid molecules, SECP proteins, and anti-SECP antibodies (also referred to herein as "active compounds") of the invention, and derivatives, fragments, analogs and homologs thereof, can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically-acceptable carrier. As used herein, "pharmaceutically-acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Suitable carriers are described in the most recent edition of Remington's Pharmaceutical Sciences, a standard reference text in the field, which is incorporated herein by reference. Preferred examples of such carriers or diluents include, but are not limited to, water, saline, finger's solutions, dextrose solution, and 5% human serum albumin. Liposomes and other non-aqueous (*i.e.*, lipophilic) vehicles such as fixed oils may also be used. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with

the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (*i.e.*, topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid (EDTA); buffers such as acetates, citrates or phosphates, and agents for the adjustment of tonicity such as sodium chloride or dextrose. The pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, N.J.) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringeability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (*e.g.*, a SECP protein or anti-SECP antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle that contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, methods of preparation are vacuum drying and freeze-drying that yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (*see, e.g.,* U.S. Patent No. 5,328,470) or by stereotactic injection (*see, e.g.*, Chen, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91: 3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, *e.g.*, retroviral vectors, the pharmaceutical preparation can include one or more cells that produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

Screening and Detection Methods

The nucleic acid molecules, proteins, protein homologues, and antibodies described herein can be used in one or more of the following methods: (A) screening assays; (B) detection assays (*e.g.*, chromosomal mapping, cell and tissue typing, forensic biology), (C) predictive medicine (*e.g.*, diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenomics); and (D) methods of treatment (*e.g.*, therapeutic and prophylactic).

The isolated nucleic acid molecules of the present invention can be used to express SECP protein (e.g., via a recombinant expression vector in a host cell in gene therapy applications), to detect SECP mRNA (e.g., in a biological sample) or a genetic lesion in an SECP gene, and to modulate SECP activity, as described further below. In addition, the SECP proteins can be used to screen drugs or compounds that modulate the SECP protein activity or expression as well as to treat disorders characterized by insufficient or excessive production of SECP protein or production of SECP protein forms that have decreased or aberrant activity compared to SECP wild-type protein. In addition, the anti-SECP antibodies of the present invention can be used to detect and isolate SECP proteins and modulate SECP activity.

The invention further pertains to novel agents identified by the screening assays described herein and uses thereof for treatments as previously described.

Screening Assays

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, *i.e.*, candidate or test compounds or agents (*e.g.*, peptides, peptidomimetics, small molecules or other drugs) that bind to SECP proteins or have a stimulatory or inhibitory effect on, *e.g.*, SECP protein expression or SECP protein activity. The invention also includes compounds identified in the screening assays described herein.

In one embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of the membrane-bound form of a SECP protein or polypeptide or biologically-active portion thereof. The test compounds of the invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity

chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds. *See, e.g.,* Lam, 1997. *Anticancer Drug Design* 12: 145.

A "small molecule" as used herein, is meant to refer to a composition that has a molecular weight of less than about 5 kD and most preferably less than about 4 kD. Small molecules can be, *e.g.*, nucleic acids, peptides, polypeptides, peptidomimetics, carbohydrates, lipids or other organic or inorganic molecules. Libraries of chemical and/or biological mixtures, such as fungal, bacterial, or algal extracts, are known in the art and can be screened with any of the assays of the invention.

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt, *et al.*, 1993. *Proc. Natl. Acad. Sci. U.S.A.* 90: 6909; Erb, *et al.*, 1994. *Proc. Natl. Acad. Sci. U.S.A.* 91: 11422; Zuckermann, *et al.*, 1994. *J. Med. Chem.* 37: 2678; Cho, *et al.*, 1993. *Science* 261: 1303; Carrell, *et al.*, 1994. *Angew. Chem. Int. Ed. Engl.* 33: 2059; Carell, *et al.*, 1994. *Angew. Chem. Int. Ed. Engl.* 33: 2061; and Gallop, *et al.*, 1994. *J. Med. Chem.* 37: 1233.

Libraries of compounds may be presented in solution (*e.g.*, Houghten, 1992. *Biotechniques* 13: 412-421), or on beads (Lam, 1991. *Nature* 354: 82-84), on chips (Fodor, 1993. *Nature* 364: 555-556), bacteria (Ladner, U.S. Patent No. 5,223,409), spores (Ladner, U.S. Patent 5,233,409), plasmids (Cull, *et al.*, 1992. *Proc. Natl. Acad. Sci. USA* 89: 1865-1869) or on phage (Scott and Smith, 1990. *Science* 249: 386-390; Devlin, 1990. *Science* 249: 404-406; Cwirla, *et al.*, 1990. *Proc. Natl. Acad. Sci. U.S.A.* 87: 6378-6382; Felici, 1991. *J. Mol. Biol.* 222: 301-310; Ladner, U.S. Patent No. 5,233,409.).

In one embodiment, an assay is a cell-based assay in which a cell which expresses a membrane-bound form of SECP protein, or a biologically-active portion thereof, on the cell surface is contacted with a test compound and the ability of the test compound to bind to a SECP protein determined. The cell, for example, can of mammalian origin or a yeast cell. Determining the ability of the test compound to bind to the SECP protein can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the SECP protein or biologically-active portion thereof can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with ^{125}I, ^{35}S, ^{14}C, or ^{3}H, either directly or indirectly, and the radioisotope detected by direct counting of radioemission or by scintillation counting. Alternatively, test compounds can

be enzymatically-labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product. In one embodiment, the assay comprises contacting a cell which expresses a membrane-bound form of SECP protein, or a biologically-active portion thereof, on the cell surface with a known compound which binds SECP to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a SECP protein, wherein determining the ability of the test compound to interact with a SECP protein comprises determining the ability of the test compound to preferentially bind to SECP protein or a biologically-active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a membrane-bound form of SECP protein, or a biologically-active portion thereof, on the cell surface with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the SECP protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of SECP or a biologically-active portion thereof can be accomplished, for example, by determining the ability of the SECP protein to bind to or interact with a SECP target molecule. As used herein, a "target molecule" is a molecule with which a SECP protein binds or interacts in nature, for example, a molecule on the surface of a cell which expresses a SECP interacting protein, a molecule on the surface of a second cell, a molecule in the extracellular *milieu*, a molecule associated with the internal surface of a cell membrane or a cytoplasmic molecule. An SECP target molecule can be a non-SECP molecule or a SECP protein or polypeptide of the invention. In one embodiment, a SECP target molecule is a component of a signal transduction pathway that facilitates transduction of an extracellular signal (*e.g.* a signal generated by binding of a compound to a membrane-bound SECP molecule) through the cell membrane and into the cell. The target, for example, can be a second intercellular protein that has catalytic activity or a protein that facilitates the association of downstream signaling molecules with SECP.

Determining the ability of the SECP protein to bind to or interact with a SECP target molecule can be accomplished by one of the methods described above for determining direct binding. In one embodiment, determining the ability of the SECP protein to bind to or interact with a SECP target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (*i.e.* intracellular Ca^{2+}, diacylglycerol, IP_3, etc.), detecting catalytic/enzymatic activity of the target an appropriate substrate, detecting the

induction of a reporter gene (comprising a SECP-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, *e.g.*, luciferase), or detecting a cellular response, for example, cell survival, cellular differentiation, or cell proliferation.

In yet another embodiment, an assay of the invention is a cell-free assay comprising contacting a SECP protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to bind to the SECP protein or biologically-active portion thereof. Binding of the test compound to the SECP protein can be determined either directly or indirectly as described above. In one such embodiment, the assay comprises contacting the SECP protein or biologically-active portion thereof with a known compound which binds SECP to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a SECP protein, wherein determining the ability of the test compound to interact with a SECP protein comprises determining the ability of the test compound to preferentially bind to SECP or biologically-active portion thereof as compared to the known compound.

In still another embodiment, an assay is a cell-free assay comprising contacting SECP protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to modulate (*e.g.* stimulate or inhibit) the activity of the SECP protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of SECP can be accomplished, for example, by determining the ability of the SECP protein to bind to a SECP target molecule by one of the methods described above for determining direct binding. In an alternative embodiment, determining the ability of the test compound to modulate the activity of SECP protein can be accomplished by determining the ability of the SECP protein further modulate a SECP target molecule. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as described, *supra*.

In yet another embodiment, the cell-free assay comprises contacting the SECP protein or biologically-active portion thereof with a known compound which binds SECP protein to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a SECP protein, wherein determining the ability of the test compound to interact with a SECP protein comprises determining the ability of the SECP protein to preferentially bind to or modulate the activity of a SECP target molecule.

The cell-free assays of the invention are amenable to use of both the soluble form or the membrane-bound form of SECP protein. In the case of cell-free assays comprising the membrane-bound form of SECP protein, it may be desirable to utilize a solubilizing agent such that the membrane-bound form of SECP protein is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®, Isotridecypoly(ethylene glycol ether)$_n$, N-dodecyl--N,N-dimethyl-3-ammonio-1-propane sulfonate, 3-(3-cholamidopropyl) dimethylamminiol-1-propane sulfonate (CHAPS), or 3-(3-cholamidopropyl)dimethylamminiol-2-hydroxy-1-propane sulfonate (CHAPSO).

In more than one embodiment of the above assay methods of the invention, it may be desirable to immobilize either SECP protein or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to SECP protein, or interaction of SECP protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided that adds a domain that allows one or both of the proteins to be bound to a matrix. For example, GST-SECP fusion proteins or GST-target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, that are then combined with the test compound or the test compound and either the non-adsorbed target protein or SECP protein, and the mixture is incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described, *supra*. Alternatively, the complexes can be dissociated from the matrix, and the level of SECP protein binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either the SECP protein or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated SECP protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well-known within the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, Ill.), and immobilized

in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with SECP protein or target molecules, but which do not interfere with binding of the SECP protein to its target molecule, can be derivatized to the wells of the plate, and unbound target or SECP protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the SECP protein or target molecule, as well as enzyme-linked assays that rely on detecting an enzymatic activity associated with the SECP protein or target molecule.

In another embodiment, modulators of SECP protein expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of SECP mRNA or protein in the cell is determined. The level of expression of SECP mRNA or protein in the presence of the candidate compound is compared to the level of expression of SECP mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of SECP mRNA or protein expression based upon this comparison. For example, when expression of SECP mRNA or protein is greater (*i.e.*, statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of SECP mRNA or protein expression. Alternatively, when expression of SECP mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of SECP mRNA or protein expression. The level of SECP mRNA or protein expression in the cells can be determined by methods described herein for detecting SECP mRNA or protein.

In yet another aspect of the invention, the SECP proteins can be used as "bait proteins" in a two-hybrid assay or three hybrid assay (*see, e.g.*, U.S. Patent No. 5,283,317; Zervos, *et al.*, 1993. *Cell* 72: 223-232; Madura, *et al.*, 1993. *J. Biol. Chem.* 268: 12046-12054; Bartel, *et al.*, 1993. *Biotechniques* 14: 920-924; Iwabuchi, *et al.*, 1993. *Oncogene* 8: 1693-1696; and Brent WO 94/10300), to identify other proteins that bind to or interact with SECP ("SECP-binding proteins" or "SECP-bp") and modulate SECP activity. Such SECP-binding proteins are also likely to be involved in the propagation of signals by the SECP proteins as, for example, upstream or downstream elements of the SECP pathway.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for SECP is fused to a gene

encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a SECP-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (*e.g.*, LacZ) that is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene that encodes the protein which interacts with SECP.

The invention further pertains to novel agents identified by the aforementioned screening assays and uses thereof for treatments as described herein.

Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. By way of example, and not of limitation, these sequences can be used to: (*i*) map their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (*ii*) identify an individual from a minute biological sample (tissue typing); and (*iii*) aid in forensic identification of a biological sample. Some of these applications are described in the subsections below.

Chromosome Mapping

Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. This process is called chromosome mapping. Accordingly, portions or fragments of the SECP sequences shown in SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and 56, or fragments or derivatives thereof, can be used to map the location of the SECP genes, respectively, on a chromosome. The mapping of the SECP sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Briefly, SECP genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the SECP sequences. Computer analysis of the SECP, sequences can be used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those

hybrids containing the human gene corresponding to the SECP sequences will yield an amplified fragment.

Somatic cell hybrids are prepared by fusing somatic cells from different mammals (*e.g.*, human and mouse cells). As hybrids of human and mouse cells grow and divide, they gradually lose human chromosomes in random order, but retain the mouse chromosomes. By using media in which mouse cells cannot grow, because they lack a particular enzyme, but in which human cells can, the one human chromosome that contains the gene encoding the needed enzyme will be retained. By using various media, panels of hybrid cell lines can be established. Each cell line in a panel contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, allowing easy mapping of individual genes to specific human chromosomes. *See, e.g.*, D'Eustachio, *et al.*, 1983. *Science* 220: 919-924. Somatic cell hybrids containing only fragments of human chromosomes can also be produced by using human chromosomes with translocations and deletions.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the SECP sequences to design oligonucleotide primers, sub-localization can be achieved with panels of fragments from specific chromosomes.

Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical like colcemid that disrupts the mitotic spindle. The chromosomes can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases, will suffice to get good results at a reasonable amount of time. For a review of this technique, *see,* Verma, *et al.*, HUMAN CHROMOSOMES: A MANUAL OF BASIC TECHNIQUES (Pergamon Press, New York 1988).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to non-coding regions of

the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, *e.g.,* in McKusick, MENDELIAN INHERITANCE IN MAN, available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, *e.g.,* Egeland, *et al.*, 1987. *Nature*, 325: 783-787.

Additionally, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the SECP gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

Tissue Typing

The SECP sequences of the invention can also be used to identify individuals from minute biological samples. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. The sequences of the invention are useful as additional DNA markers for RFLP ("restriction fragment length polymorphisms," as described in U.S. Patent No. 5,272,057).

Furthermore, the sequences of the invention can be used to provide an alternative technique that determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the SECP sequences described herein can be used to prepare two PCR primers from the 5'- and 3'-termini of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA

sequences due to allelic differences. The sequences of the invention can be used to obtain such identification sequences from individuals and from tissue. The SECP sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the non-coding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Much of the allelic variation is due to single nucleotide polymorphisms (SNPs), which include restriction fragment length polymorphisms (RFLPs).

Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the non-coding regions, fewer sequences are necessary to differentiate individuals. The non-coding sequences can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers that each yield a non-coding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and 56 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

Predictive Medicine

The invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the invention relates to diagnostic assays for determining SECP protein and/or nucleic acid expression as well as SECP activity, in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant SECP expression or activity. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with SECP protein, nucleic acid expression or activity. For example, mutations in a SECP gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with SECP protein, nucleic acid expression or activity.

Another aspect of the invention provides methods for determining SECP protein, nucleic acid expression or SECP activity in an individual to thereby select appropriate therapeutic or prophylactic agents for that individual (referred to herein as "pharmacogenomics").

Pharmacogenomics allows for the selection of agents (*e.g.*, drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (*e.g.*, the genotype of the individual examined to determine the ability of the individual to respond to a particular agent.)

Yet another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of SECP in clinical trials.

Use of Partial SECP Sequences in Forensic Biology

DNA-based identification techniques can also be used in forensic biology. Forensic biology is a scientific field employing genetic typing of biological evidence found at a crime scene as a means for positively identifying, *e.g.*, a perpetrator of a crime. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues (*e.g.*, hair or skin, or body fluids, *e.g.*, blood, saliva, or semen found at a crime scene). The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

The sequences of the invention can be used to provide polynucleotide reagents, *e.g.*, PCR primers, targeted to specific loci in the human genome, that can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (*i.e.* another DNA sequence that is unique to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to non-coding regions of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and 56 are particularly appropriate for this use as greater numbers of polymorphisms occur in the non-coding regions, making it easier to differentiate individuals using this technique. Examples of polynucleotide reagents include the SECP sequences or portions thereof, *e.g.*, fragments derived from the non-coding regions of one or more of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and 56 having a length of at least 20 bases, preferably at least 30 bases.

The SECP sequences described herein can further be used to provide polynucleotide reagents, *e.g.*, labeled or label-able probes that can be used, for example, in an *in situ* hybridization technique, to identify a specific tissue (*e.g.*, brain tissue, etc). This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such SECP probes can be used to identify tissue by species and/or by organ type.

In a similar fashion, these reagents, *e.g.*, SECP primers or probes can be used to screen tissue culture for contamination (*i.e.*, screen for the presence of a mixture of different types of cells in a culture).

Predictive Medicine

The invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the invention relates to diagnostic assays for determining SECP protein and/or nucleic acid expression as well as SECP activity, in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant SECP expression or activity. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with SECP protein, nucleic acid expression or activity. For example, mutations in a SECP gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with SECP protein, nucleic acid expression, or biological activity.

Another aspect of the invention provides methods for determining SECP protein, nucleic acid expression or activity in an individual to thereby select appropriate therapeutic or prophylactic agents for that individual (referred to herein as "pharmacogenomics"). Pharmacogenomics allows for the selection of agents (*e.g.*, drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (*e.g.*, the genotype of the individual examined to determine the ability of the individual to respond to a particular agent.)

Yet another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of SECP in clinical trials.

These and various other agents are described in further detail in the following sections.

Diagnostic Assays

An exemplary method for detecting the presence or absence of SECP in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting SECP protein or nucleic acid (*e.g.*, mRNA, genomic DNA) that encodes SECP protein such that the presence of SECP is detected in

the biological sample. An agent for detecting SECP mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to SECP mRNA or genomic DNA. The nucleic acid probe can be, for example, a full-length SECP nucleic acid, such as the nucleic acid of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 40, 42, 44, 46, 48, 50, 52, 54 and 56 or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to SECP mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays of the invention are described herein.

An agent for detecting SECP protein is an antibody capable of binding to SECP protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, F_{ab} or $F_{(ab)2}$) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently-labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently-labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect SECP mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of SECP mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of SECP protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. *In vitro* techniques for detection of SECP genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of SECP protein include introducing into a subject a labeled anti-SECP antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable

of detecting SECP protein, mRNA, or genomic DNA, such that the presence of SECP protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of SECP protein, mRNA or genomic DNA in the control sample with the presence of SECP protein, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of SECP in a biological sample. For example, the kit can comprise: a labeled compound or agent capable of detecting SECP protein or mRNA in a biological sample; means for determining the amount of SECP in the sample; and means for comparing the amount of SECP in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect SECP protein or nucleic acid.

Prognostic Assays

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant SECP expression or activity. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with SECP protein, nucleic acid expression or activity. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing a disease or disorder. Thus, the invention provides a method for identifying a disease or disorder associated with aberrant SECP expression or activity in which a test sample is obtained from a subject and SECP protein or nucleic acid (*e.g.*, mRNA, genomic DNA) is detected, wherein the presence of SECP protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant SECP expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (*e.g.*, serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant SECP expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a disorder. Thus, the invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant SECP expression or activity in which a test sample is obtained and SECP protein or nucleic acid is detected (*e.g.*, wherein the presence of SECP

protein or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant SECP expression or activity).

The methods of the invention can also be used to detect genetic lesions in a SECP gene, thereby determining if a subject with the lesioned gene is at risk for a disorder characterized by aberrant cell proliferation and/or differentiation. In various embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion characterized by at least one of an alteration affecting the integrity of a gene encoding a SECP-protein, or the mis-expression of the SECP gene. For example, such genetic lesions can be detected by ascertaining the existence of at least one of: (*i*) a deletion of one or more nucleotides from a SECP gene; (*ii*) an addition of one or more nucleotides to a SECP gene; (*iii*) a substitution of one or more nucleotides of a SECP gene, (*iv*) a chromosomal rearrangement of a SECP gene; (*v*) an alteration in the level of a messenger RNA transcript of a SECP gene, (*vi*) aberrant modification of a SECP gene, such as of the methylation pattern of the genomic DNA, (*vii*) the presence of a non-wild-type splicing pattern of a messenger RNA transcript of a SECP gene, (*viii*) a non-wild-type level of a SECP protein, (*ix*) allelic loss of a SECP gene, and (*x*) inappropriate post-translational modification of a SECP protein. As described herein, there are a large number of assay techniques known in the art which can be used for detecting lesions in a SECP gene. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (*see, e.g.*, U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (*see, e.g.*, Landegran, *et al.*, 1988. *Science* 241: 1077-1080; and Nakazawa, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91: 360-364), the latter of which can be particularly useful for detecting point mutations in the SECP-gene (*see,* Abravaya, *et al.*, 1995. *Nucl. Acids Res.* 23: 675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (*e.g.*, genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers that specifically hybridize to a SECP gene under conditions such that hybridization and amplification of the SECP gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be

desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (*see,* Guatelli, *et al.*, 1990. *Proc. Natl. Acad. Sci. USA* 87: 1874-1878), transcriptional amplification system (*see,* Kwoh, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA* 86: 1173-1177); Qβ Replicase (*see,* Lizardi, *et al*, 1988. *BioTechnology* 6: 1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in a SECP gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (*see, e.g.,* U.S. Patent No. 5,493,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in SECP can be identified by hybridizing a sample and control nucleic acids, *e.g.*, DNA or RNA, to high-density arrays containing hundreds or thousands of oligonucleotides probes. *See, e.g.,* Cronin, *et al.*, 1996. *Human Mutation* 7: 244-255; Kozal, *et al.*, 1996. *Nat. Med.* 2: 753-759. For example, genetic mutations in SECP can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, *et al.*, *supra.* Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the SECP gene and detect mutations by comparing the sequence of the sample SECP with the corresponding wild-type (control) sequence. Examples of sequencing

reactions include those based on techniques developed by Maxim and Gilbert, 1977. *Proc. Natl. Acad. Sci. USA* 74: 560 or Sanger, 1977. *Proc. Natl. Acad. Sci. USA* 74: 5463. It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays (*see, e.g.,* Naeve, *et al.*, 1995. *Biotechniques* 19: 448), including sequencing by mass spectrometry (see, *e.g.*, PCT International Publication No. WO 94/16101; Cohen, *et al.*, 1996. *Adv. Chromatography* 36: 127-162; and Griffin, *et al.*, 1993. *Appl. Biochem. Biotechnol.* 38: 147-159).

Other methods for detecting mutations in the SECP gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes. *See, e.g.,* Myers, *et al.*, 1985. *Science* 230: 1242. In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type SECP sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent that cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S_1 nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. *See, e.g.,* Cotton, *et al.*, 1988. *Proc. Natl. Acad. Sci. USA* 85: 4397; Saleeba, *et al.*, 1992. *Methods Enzymol.* 217: 286-295. In an embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in SECP cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches. *See, e.g.,* Hsu, *et al.*, 1994. *Carcinogenesis* 15: 1657-1662. According to an exemplary embodiment, a probe based on a SECP sequence, *e.g.*, a wild-type SECP sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. *See, e.g.,* U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in SECP genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids. *See, e.g.,* Orita, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA*: 86: 2766; Cotton, 1993. *Mutat. Res.* 285: 125-144; Hayashi, 1992. *Genet. Anal. Tech. Appl.* 9: 73-79. Single-stranded DNA fragments of sample and control SECP nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In one embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility. *See, e.g.,* Keen, *et al.*, 1991. *Trends Genet.* 7: 5.

In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE). *See, e.g.,* Myers, *et al.,* 1985. *Nature* 313: 495. When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA. *See, e.g.,* Rosenbaum and Reissner, 1987. *Biophys. Chem.* 265: 12753.

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions that permit hybridization only if a perfect match is found. *See, e.g.,* Saiki, *et al.*, 1986. *Nature* 324: 163; Saiki, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA* 86: 6230. Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology that depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization; *see, e.g.,* Gibbs, *et al.*, 1989. *Nucl.*

Acids Res. 17: 2437-2448) or at the extreme 3'-terminus of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (*see, e.g.,* Prossner, 1993. *Tibtech.* 11: 238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection. *See, e.g.,* Gasparini, *et al.,* 1992. *Mol. Cell Probes* 6: 1. It is anticipated that in certain embodiments amplification may also be performed using *Taq* ligase for amplification. *See, e.g.,* Barany, 1991. *Proc. Natl. Acad. Sci. USA* 88: 189. In such cases, ligation will occur only if there is a perfect match at the 3'-terminus of the 5' sequence, making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.*, in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a SECP gene.

Furthermore, any cell type or tissue, preferably peripheral blood leukocytes, in which SECP is expressed may be utilized in the prognostic assays described herein. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

Pharmacogenomics

Agents, or modulators that have a stimulatory or inhibitory effect on SECP activity (*e.g.*, SECP gene expression), as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders (*e.g.*, cancer or immune disorders associated with aberrant SECP activity. In conjunction with such treatment, the pharmacogenomics (*i.e.*, the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) of the individual may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the selection of effective agents (*e.g.*, drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such pharmacogenomics can further be used to determine appropriate dosages and therapeutic regimens. Accordingly, the activity of SECP protein, expression of SECP nucleic acid, or mutation content of SECP genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See *e.g.*, Eichelbaum, 1996. *Clin. Exp. Pharmacol. Physiol.* 23: 983-985; Linder, 1997. *Clin. Chem.*, 43: 254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase (G6PD) deficiency is a common inherited enzymopathy in which the main clinical complication is hemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. At the other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Thus, the activity of SECP protein, expression of SECP nucleic acid, or mutation content of SECP genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance

therapeutic or prophylactic efficiency when treating a subject with a SECP modulator, such as a modulator identified by one of the exemplary screening assays described herein.

Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of SECP (*e.g.*, the ability to modulate aberrant cell proliferation and/or differentiation) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase SECP gene expression, protein levels, or upregulate SECP activity, can be monitored in clinical trails of subjects exhibiting decreased SECP gene expression, protein levels, or down-regulated SECP activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease SECP gene expression, protein levels, or down-regulate SECP activity, can be monitored in clinical trails of subjects exhibiting increased SECP gene expression, protein levels, or up-regulated SECP activity. In such clinical trials, the expression or activity of SECP and, preferably, other genes that have been implicated in, for example, a cellular proliferation or immune disorder can be used as a "read out" or markers of the immune responsiveness of a particular cell.

By way of example, and not of limitation, genes, including SECP, that are modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule) that modulates SECP activity (*e.g.*, identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on cellular proliferation disorders, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of SECP and other genes implicated in the disorder. The levels of gene expression (*i.e.*, a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of SECP or other genes. In this manner, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

In one embodiment, the invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (*e.g.*, an agonist, antagonist, protein, peptide, peptidomimetic, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (*i*) obtaining a pre-administration sample from a

subject prior to administration of the agent; (*ii*) detecting the level of expression of a SECP protein, mRNA, or genomic DNA in the pre-administration sample; (*iii*) obtaining one or more post-administration samples from the subject; (*iv*) detecting the level of expression or activity of the SECP protein, mRNA, or genomic DNA in the post-administration samples; (*v*) comparing the level of expression or activity of the SECP protein, mRNA, or genomic DNA in the pre-administration sample with the SECP protein, mRNA, or genomic DNA in the post administration sample or samples; and (*vi*) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of SECP to higher levels than detected, *i.e.*, to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of SECP to lower levels than detected, *i.e.*, to decrease the effectiveness of the agent.

Methods of Treatment

The invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant SECP expression or activity. These methods of treatment will be discussed more fully, below.

Disease and Disorders

Diseases and disorders that are characterized by increased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that antagonize (*i.e.*, reduce or inhibit) activity. Therapeutics that antagonize activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to: (*i*) an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; (*ii*) antibodies to an aforementioned peptide; (*iii*) nucleic acids encoding an aforementioned peptide; (*iv*) administration of antisense nucleic acid and nucleic acids that are "dysfunctional" (*i.e.*, due to a heterologous insertion within the coding sequences of coding sequences to an aforementioned peptide) that are utilized to "knockout" endoggenous function of an aforementioned peptide by homologous recombination (*see*, *e.g.*, Capecchi, 1989. *Science* 244: 1288-1292); or (*v*) modulators (*i.e.*, inhibitors, agonists and antagonists, including additional peptide mimetic of the invention or antibodies specific to a peptide of the invention) that alter the interaction between an aforementioned peptide and its binding partner.

Diseases and disorders that are characterized by decreased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with

Therapeutics that increase (*i.e.*, are agonists to) activity. Therapeutics that upregulate activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to, an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; or an agonist that increases bioavailability.

Increased or decreased levels can be readily detected by quantifying peptide and/or RNA, by obtaining a patient tissue sample (*e.g.*, from biopsy tissue) and assaying it *in vitro* for RNA or peptide levels, structure and/or activity of the expressed peptides (or mRNAs of an aforementioned peptide). Methods that are well-known within the art include, but are not limited to, immunoassays (*e.g.*, by Western blot analysis, immunoprecipitation followed by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis, immunocytochemistry, etc.) and/or hybridization assays to detect expression of mRNAs (*e.g.*, Northern assays, dot blots, *in situ* hybridization, and the like).

Prophylactic Methods

In one aspect, the invention provides a method for preventing, in a subject, a disease or condition associated with an aberrant SECP expression or activity, by administering to the subject an agent that modulates SECP expression or at least one SECP activity. Subjects at risk for a disease that is caused or contributed to by aberrant SECP expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the SECP aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending upon the type of SECP aberrancy, for example, a SECP agonist or SECP antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein.

Therapeutic Methods

Another aspect of the invention pertains to methods of modulating SECP expression or activity for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of SECP protein activity associated with the cell. An agent that modulates SECP protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of a SECP protein, a peptide, a SECP peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more SECP protein activity. Examples of such stimulatory agents include active SECP protein and a nucleic acid molecule encoding SECP that has been

introduced into the cell. In another embodiment, the agent inhibits one or more SECP protein activity. Examples of such inhibitory agents include antisense SECP nucleic acid molecules and anti-SECP antibodies. These modulatory methods can be performed *in vitro* (*e.g.*, by culturing the cell with the agent) or, alternatively, *in vivo* (*e.g.*, by administering the agent to a subject). As such, the invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of a SECP protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (*e.g.*, an agent identified by a screening assay described herein), or combination of agents that modulates (*e.g.*, up-regulates or down-regulates) SECP expression or activity. In another embodiment, the method involves administering a SECP protein or nucleic acid molecule as therapy to compensate for reduced or aberrant SECP expression or activity.

Stimulation of SECP activity is desirable in situations in which SECP is abnormally down-regulated and/or in which increased SECP activity is likely to have a beneficial effect. One example of such a situation is where a subject has a disorder characterized by aberrant cell proliferation and/or differentiation (*e.g.*, cancer or immune associated disorders). Another example of such a situation is where the subject has a gestational disease (*e.g.*, pre-clampsia).

Determination of the Biological Effect of the Therapeutic

In various embodiments of the invention, suitable *in vitro* or *in vivo* assays are performed to determine the effect of a specific Therapeutic and whether its administration is indicated for treatment of the affected tissue.

In various specific embodiments, *in vitro* assays may be performed with representative cells of the type(s) involved in the patient's disorder, to determine if a given Therapeutic exerts the desired effect upon the cell type(s). Compounds for use in therapy may be tested in suitable animal model systems including, but not limited to rats, mice, chicken, cows, monkeys, rabbits, and the like, prior to testing in human subjects. Similarly, for *in vivo* testing, any of the animal model system known in the art may be used prior to administration to human subjects.

Prophylactic and Therapeutic Uses of the Compositions of the Invention

The SECP nucleic acids and proteins of the invention may be useful in a variety of potential prophylactic and therapeutic applications. By way of a non-limiting example, a cDNA encoding the SECP protein of the invention may be useful in gene therapy, and the protein may be useful when administered to a subject in need thereof.

Both the novel nucleic acids encoding the SECP proteins, and the SECP proteins of the invention, or fragments thereof, may also be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. These materials are further useful in the generation of antibodies which immunospecifically-bind to the novel substances of the invention for use in therapeutic or diagnostic methods.

The invention will be further illustrated in the following non-limiting examples.

Example 1: Radiation Hybrid Mapping Provides the Chromosomal Location of SECP 2 (Clone 11618130.0.27)

Radiation hybrid mapping using human chromosome markers was carried out to determine the chromosomal location of a SECP2 nuclei acid of the invention. The procedure used to obtain these results is described generally in Steen, *et al.,* 1999. A High-Density Integrated Genetic Linkage and Radiation Hybrid Map of the Laboratory Rat, *Genome Res.* 9: AP1-AP8 (Published Online on May 21, 1999). A panel of 93 cell clones containing randomized radiation-induced human chromosomal fragments was then screened in 96 well plates using PCR primers designed to identify the sought clones in a unique fashion. Clone 11618130.0.27, a SECP2 nucleic acid was located on chromosome 16 at a map distance of 26.0 cR from marker WI-3768 and -70.5 cR from marker TIGR-A002K05.

Example 2: Molecular Cloning of Clone 11618130

Oligonucleotide PCR primers were designed to amplify a DNA segment coding for the full length open reading frame of clone 11618130. The forward primer included a Bgl II restriction site and the consensus Kozak sequence CCACC. The reverse primer contained an in-frame XhoI restriction site. Both primers contained a CTCGTC 5'-terminus clamp. The nucleotide sequences of the primers were:

11618130 Forward Primer:

CTCGTCAGATCTCCACCATGAGTGATGAGGACAGCTGTGTAG (SEQ ID NO:19)

11618130 Reverse Primer:

CTCGTCCTCGAGGCAGCTGGTTGGTTGGCTTATGTTG (SEQ ID NO:20)

The PCR reactions included: 5 ng human fetal brain cDNA template; 1 μM of each of the 11618130 Forward and 11618130 Reverse primers; 5 μM dNTP (Clontech Laboratories; Palo Alto, CA) and 1 μl of 50x Advantage-HF 2 polymerase (Clontech Laboratories; Palo Alto, CA) in 50 μl total reaction volume. The following PCR conditions were used:

a) 96°C 3 minutes

b) 96°C 30 seconds denaturation

c) 70°C 30 seconds, primer annealing. This temperature was gradually decreased by 1°C/cycle

d) 72°C 1 minute extension.

Repeat steps b-d a total of 10-times

e) 96°C 30 seconds denaturation

f) 60°C 30 seconds annealing

g) 72°C 1 minute extension

Repeat steps e-g a total of 25-times

h) 72°C 5 minutes final extension

A single, amplified product of approximately 800 bp was detected by agarose gel electrophoresis. The PCR amplification product was then isolated by the QIAEX II® Gel Extraction System (QIAGEN, Inc; Valencia, CA) in a final volume of 20 µl.

A total of 10 µl of the isolated fragment was digested with Bgl II and XhoI restriction enzymes, and ligated into the BamHI- and XhoI-digested mammalian expression vector pCDNA3.1 V5His (Invitrogen; Carlsbad, CA.). The construct was sequenced, and the cloned insert was verified as a sequence identical to the ORF coding for the full length 11618130. The construct was designated pcDNA3.1-11618130-S178-2.

Example 3: Expression of 11618130 In Human Embryonic Kidney 293 Cells

The vector pcDNA3.1-11618130-S178-2 described in Example 2 was subsequently transfected into human embryonic kidney 293 cells (ATCC No. CRL-1573; Manassas, VA) using the LipofectaminePlus Reagent following the manufacturer's instructions (Gibco/BRL/Life Technologies; Rockville, MD) The cell pellet and supernatant were harvested 72 hours after transfection, and examined for 11618130 expression by use of SDS-PAGE under reducing conditions and Western blotting with an anti-V5 antibody. FIG. 12 shows that 11618130 was expressed as a protein having an apparent molecular weight (Mr) of approximately 34 kilo Daltons (kDa) which was intracellularly expressed in the 293 cells. These experimental results were consistent with the predicted molecular weight of 28043 Daltons for the protein of clone 11618130.0.27 and with the predicted localization of the protein intracellularly in the microbody (peroxisome). A second band of approximately 54 kDa was also found, which may represent a non-reducible dimer of this protein.

Example 4: Preparation of Mammalian Expression Vector pSecV5His

The oligonucleotide primers, pSec-V5-His Forward and pSec-V5-His Reverse, were generated to amplify a fragment from the pcDNA3.1-V5His (Invitrogen; Carlsbad, CA) expression vector that includes V5 and His6. The nucleotide sequences of these primers were:

pSec-V5-His Forward Primer:

CTCGTCCTCGAGGGTAAGCCTATCCCTAAC (SEQ ID NO:21)

pSec-V5-His Reverse Primer:

CTCGTCGGGCCCCTGATCAGCGGGTTTAAAC (SEQ ID NO:22)

The PCR product was digested with XhoI and ApaI, and ligated into the XhoI/ApaI-digested pSecTag2 B vector harboring an Ig kappa leader sequence (Invitrogen; Carlsbad, CA). The correct structure of the resulting vector (designated pSecV5His), including an in-frame Ig-kappa leader and V5-His6, was verified by DNA sequence analysis. The pSecV5His vector included an in-frame Ig kappa leader, a site for insertion of a clone of interest, V5 and His6, which allows heterologous protein expression and secretion by fusing any protein to the Ig kappa chain signal peptide. Detection and purification of the expressed protein was aided by the presence of the V5 epitope tag and 6x His tag at the carboxyl-terminus (Invitrogen; Carlsbad, CA).

Example 5: Molecular Cloning of 16406477

Oligonucleotide PCR primers were designed to amplify a DNA segment encoding for the mature form of clone 16406477 from amino acid residues 38 to 385, recognition of the signal sequence predicted for this polypeptide. The forward primer contained an in-frame BamHI restriction site and the reverse primer contained an in-frame XhoI restriction site. Both primers contained the CTCGTC 5' clamp. The sequences of the primers were as follows:

16406477 Forward Primer:

CTCGTCGGATCCTGGGGCGCAGGGGAAGCCCCGGG (SEQ ID NO:23)

16406477 Reverse Primer:

CTCGTCCTCGAGGAGGGCAGCAAGGAGGCTGAGGGGCAG (SEQ ID NO:24)

The PCR reactions contained: 5 ng human fetal brain cDNA template; 1 μM of each of the 16406477 Forward and 16406477 Reverse Primers; 5 μM dNTP (Clontech Laboratories; Palo Alto, CA) and 1 μl of 50x Advantage-HF 2 polymerase (Clontech Laboratories; Palo Alto,

CA) in a 50 µl total reaction volume. PCR was then conducted using reaction conditions identical to those previously described in Example 2.

A single, amplified product of approximately 1 Kbp was detected by agarose gel electrophoresis. The product was then isolated by QIAEX II® Gel Extraction System (QUIAGEN, Inc; Valencia, CA) in a total reaction volume of 20 µl.

A total of 10 µl of the isolated fragment was digested with BamHI and XhoI restriction enzymes, and ligated into the pSecV5-His mammalian expression vector (*see,* Example 4) which had been previously-digested with BamHI and XhoI. The construct was sequenced, and the cloned insert was verified as possessing a sequence identical to that of the ORF coding for the mature fragment of clone 16406477. The construct was subsequently designated pSecV5His-16406477-S196-A.

Example 6: Expression of 16406477 in Human Embryonic Kidney 293 Cells

The pSecV5His-16406477-S196-A construct (*see,* Example 5) was subsequently transfected into 293 cells (ATCC No. CRL-1573; Manassas, VA) using the LipofectaminePlus Reagent following the manufacturer's instructions (Gibco/BRL/Life Technologies). The cell pellet and supernatant were harvested 72 hours after transfection, and examined for 16406477 expression by use of SDS-PAGE under reducing conditions and Western blotting with an anti-V5 antibody. FIG. 13 demonstrates that 16406477 is expressed as a protein having an apparent molecular weight (Mr) of approximately 45 kDa which is retained intracellularly in the 293 cells. The Mr value which was found upon expression of the clone is consistent with the predicted molecular weight of 43087 Daltons.

Example 7: Quantitative Tissue Expression Analysis of Clones of the Invention

The Quantitative Expression Analysis of several clones of the invention was preformed in 41 normal and 55 tumor samples (*see,* FIG. 14) by real-time quantitative PCR (TAQMAN®) by use of a Perkin-Elmer Biosystems ABI PRISM® 7700 Sequence Detection System. The following abbreviations are used in FIG. 14:

ca. = carcinoma,
* = established from metastasis,
met = metastasis,
s cell var= small cell variant,
non-s = non-sm =non-small,
squam = squamous,
pl. eff = pl effusion = pleural effusion,

glio = glioma,
astro = astrocytoma, and
neuro = neuroblastoma.

Initially, 96 RNA samples were normalized to β-actin and GAPDH. RNA (~50 ng total or ~1 ng poly(A)+) was converted to cDNA using the TAQMAN® Reverse Transcription Reagents Kit (PE Biosystems; Foster City, CA; Catalog No. N808-0234) and random hexamers according to the manufacturer's protocol. Reactions were performed in a 20 μl total volume, and incubated for 30 minutes at 48°C. cDNA (5 μl) was then transferred to a separate plate for the TAQMAN® reaction using β-actin and GAPDH TAQMAN® Assay Reagents (PE Biosystems; Catalog Nos. 4310881E and 4310884E, respectively) and TAQMAN® Universal PCR Master Mix (PE Biosystems; Catalog No. 4304447) according to the manufacturer's protocol. Reactions were performed in a 25 μl total volume using the following parameters: 2 minutes at 50°C; 10 minutes at 95°C; 15 seconds at 95°C/1 min. at 60°C (40 cycles total).

Results were recorded as CT values (*i.e.*, cycle at which a given sample crosses a threshold level of fluorescence) using a log scale, with the difference in RNA concentration between a given sample and the sample with the lowest CT value being represented as $2^{\delta CT}$. The percent relative expression is then obtained by taking the reciprocal of this RNA difference and multiplying by 100. The average CT values obtained for β-actin and GAPDH were used to normalize RNA samples. The RNA sample generating the highest CT value required no further diluting, while all other samples were diluted relative to this sample according to their β-actin /GAPDH average CT values.

Normalized RNA (5 μl) was converted to cDNA and analyzed via TAQMAN® using One Step RT-PCR Master Mix Reagents (PE Biosystems; Catalog No. 4309169) and gene-specific primers according to the manufacturer's instructions. Probes and primers were designed for each assay according to Perkin Elmer Biosystem's Primer Express Software package (Version I for Apple Computer's Macintosh Power PC) using the sequence of the respective clones as input. Default settings were used for reaction conditions and the following parameters were set before selecting primers: primer concentration = 250 nM; primer melting temperature (T_m) range = 58°-60° C; primer optimal Tm = 59° C; maximum primer difference = 2° C, probe does not posses a 5'-terminus G; probe T_m must be 10° C greater than primer T_m; and amplicon size 75 bp to 100 bp in length. The probes and primers were synthesized by Synthegen (Houston, TX). Probes were double-purified by HPLC to remove uncoupled dye and then evaluated by mass spectroscopy to verify coupling of reporter and quencher dyes to the 5'- and 3'-termini of the

probe, respectively. Their final concentrations used were - Forward and Reverse Primers = 900 nM each; and probe = 200nM.

Subsequent PCR conditions were as follows. Normalized RNA from each tissue and each cell line was spotted in each well of a 96 well PCR plate (Perkin Elmer Biosystems). PCR reaction mixes, including two probes (*i.e.,* SECP-specific and another gene-specific probe multiplexed with the SEPC-specific probe) were set up using 1x TaqMan™ PCR Master Mix for the PE Biosystems 7700, with 5 mM $MgCl_2$; dNTPs (dA, G, C, U at 1:1:1:2 ratios); 0.25 U/ml AmpliTaq Gold™ (PE Biosystems); 0.4 U/μl RNase inhibitor; and 0.25 U/μl Reverse Transcriptase. Reverse transcription was then performed at 48°C for 30 minutes, followed by amplification/PCR cycles as follows: 95°C 10 minuets, then 40 cycles of 95° C for 15 seconds, and 60°C for 1 minute.

The primer-probe sets employed in the expression analysis of each clone, and a summary of the results, are provided below. The complete experimental results are illustrated in FIG. 14. The panel of cell lines employed was identical in all cases except that samples 95 and 96 were gDNA and a melanoma UACC-257 (control), respectively, in the experiments for clone 11696905. The nucleotide sequences of the primer sets used for these clones are as follows:

Clone 11696905.0.47 Primer Set:

Ag 383 (F):	5'-GGCCTCTCCGTACCCTTCTC-3'	(SEQ ID NO:25)
Ag 383(R):	5'-AGAGGCTCTTGGCGCAGTT-3'	(SEQ ID NO:26)
Ag 383 (P):	TET-5'-ACCAGGATCACGACCTCCGCAGG-3'-TAMRA	(SEQ ID NO:27)

Primer Set Ag 383 was designed to probe for nucleotides 403-478 in SEPC 3 (clone 11696905.0.47). The results indicate that the clone was prominently expressed in normal cells such as adipose, adrenal gland, various regions of the brain, skeletal muscle, bladder, liver and fetal liver, mammary gland, placenta, prostate and testis. It was also found to be expressed at levels much higher than comparable normal cells in cancers of the kidney and lung, and expressed at levels much lower than comparable normal cells in cancers of the central nervous system (CNS) and breast. These results suggest that SEPC 3 (clone 11696905.0.47), or fragments thereof, may be useful in probing for cancer in kidney and lung, and that the nucleic acid or the protein of clone 11696905.0.47 may be a target for therapeutic agents in such cancers. These nucleic acids and proteins may be useful as therapeutic agents in treating cancers of the CNS and breast.

Clone 16406477.0.206 Primer Set:

Ag 53 (F):	5'-GCCTGGCACGGACTATGTGT-3'	(SEQ ID NO:28)
Ag 53 (R):	5'-GCCGTCAGCCTTGGAAAGT-3'	(SEQ ID NO:29)
Ag 53 (P):	TET-5'-CCATTCCCGCTGCACTGTGACG-3'-TAMRA	(SEQ ID NO:30)

SEPC 7 (clone 16406477.0.206) was found to be expressed essentially exclusively in testis cells, with a low level of expression in the hypothalamus, among the cells tested.

Clone 21433858 Primer Set:

Ag 127 (F):	5'-CCTGCCAGGATGACTGTCAATT-3'	(SEQ ID NO:31)
Ag 127 (R):	5'-TGGTCCTAACTGCACCACAGTCT-3'	(SEQ ID NO:32)
Ag 127 (P):	TET-5'-CCAGCTGGTCCAAGTTTTCTTCATGCAA-3'-TAMRA	(SEQ ID NO:33)

Probe set Ag 127 targets nucleotides 2524-2601 of SECP1 (clone 21433858). The results show that the clone is expressed principally in normal tissues such as adipose, brain, bladder, fetal and adult kidney, mammary gland, myometrium, uterus, placenta, and testis. In comparison to normal lung tissue, it is highly expressed in a small cell lung cancer, a large cell lung cancer, and a non-small cell lung cancer. Therefore, SECP1 (clone 21433858), or a fragment thereof, may be useful as a diagnostic probe for such lung cancers. The nucleic acids or proteins of SECP1 (clone 21433858) may furthermore serve as targets for the treatment of cancer in these and other tissues.

Clone 21637262.0.64 Primer Set:

Ab5(F):	5'-GTGATCCTCAGGCTGGACCA-3'	(SEQ ID NO:34)
Ab5(R):	5'-TTCTGACTGGGCTGCATCC-3'	(SEQ ID NO:35)
Ab5(P):	FAM-5'-CCAGTGTTTCCTCAGCACAGGGCC-3'-TAMRA	(SEQ ID NO:36)

Probe set Ab5 targets nucleotides 1221-1298 in SECP9 (clone 21637262.0.64). The results shown in FIG. 14 demonstrate that SECP9 (clone 21637262.0.64) is expressed in cells from normal tissues including, especially, the salivary gland and trachea, among those cells examined.

Table ??. Probe and Primer Set: Ag 815 for CG106318_01

Primers	Sequences	TM	Length	Start Position	SEQ ID NO
Forward	5'-TGTGCTCAGCACATGGTCTA-3'	59	20	1722	37
Probe	FAM-5'-ACACCTGCTCAGGGAAAACGACAGAA-	69.9	26	1760	38

	3'-TAMRA				
Reverse	5'-TCGTGCTCGTATCTGTTTCC-3'	58.9	20	1787	39

Other Embodiments

While the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

REFERENCES

1. Altshuler, D.; Hirschhorn, J. N.; Klannemark, M.; Lindgren, C. M.; Vohl, M.-C.; Nemesh, J.; Lane, C. R.; Schaffner, S. F.; Bolk, S.; Brewer, C.; Tuomi, T.; Gaudet, D.; Hudson, T. J.; Daly, M.; Groop, L.; Lander, E. S. : The common PPAR-gamma pro12ala polymorphism is associated with decreased risk of type 2 diabetes. Nature Genet. 76-80, 2000. PubMed ID : 10973253

2. Barak, Y.; Nelson, M. C.; Ong, E. S.; Jones, Y. Z.; Ruiz-Lozano, P.; Chien, K. R.; Koder, A.; Evans, R. M. : PPAR-gamma is required for placental, cardiac, and adipose tissue development. Molec. Cell 4: 585-595, 1999. PubMed ID : 10549290

3. Barroso, I.; Gurnell, M.; Crowley, V. E. F.; Agostini, M.; Schwabel, J. W.; Soos, M. A.; Masien, G. L.; Williams, T. D. M.; Lewis, H.; Schafer, A. J.; Chatterjee, V. K. K.; O'Rahilly, S. : Dominant negative mutations in human PPAR-gamma associated with severe insulin resistance, diabetes mellitus and hypertension. Nature 402: 880-883, 1999. PubMed ID : 10622252

4. Beamer, B. A.; Negri, C.; Yen, C.-J.; Gavrilova, O.; Rumberger, J. M.; Durcan, M. J.; Yarnall, D. P.; Hawkins, A. L.; Griffin, C. A.; Burns, D. K.; Roth, J.; Reitman, M.; Shuldiner, A. R. : Chromosomal localization and partial genomic structure of the human peroxisome proliferator activated receptor-gamma (hPPAR-gamma) gene. Biochem. Biophys. Res. Commun. 233: 756-759, 1997. PubMed ID : 9168928

5. Beamer, B. A.; Yen, C.-J.; Andersen, R. E.; Muller, D.; Elahi, D.; Cheskin, L. J.; Andres, R.; Roth, J.; Shuldiner, A. R. : Association of the pro12ala variant in the peroxisome proliferator-activated receptor-gamma-2 gene with obesity in two Caucasian populations. Diabetes 47: 1806-1808, 1998. PubMed ID : 9792554

6. Chawla, A.; Boisvert, W. A.; Lee, C.-H.; Laffitte, B. A.; Barak, Y.; Joseph, S. B.; Liao, D.; Nagy, L.; Edwards, P. A.; Curtiss, L. K.; Evans, R. M.; Tontonoz P. : A PPAR-gamma-LXR-ABCA1 pathway in macrophages is involved in cholesterol efflux and atherogenesis. Molec. Cell 7: 161-171, 2001. PubMed ID : 11172721

7. Deeb, S. S.; Fajas, L.; Nemoto, M.; Pihlajamaki, J.; Mykkanen, L.; Kuusisto, J.; Laakso, M.; Fujimoto, W.; Auwerx, J. : A pro12ala substitution in PPAR-gamma-2 associated with decreased receptor activity, lower body mass index and improved insulin sensitivity. Nature Genet. 20: 284-287, 1998. PubMed ID : 9806549

8. Elbrecht, A.; Chen, Y.; Cullinan, C. A.; Hayes, N.; Leibowitz, M. D.; Moller, D. E.; Berger, J.: Molecular cloning, expression and characterization of human peroxisome proliferator activated receptors gamma-1 and gamma-2. Biochem. Biophys. Res. Commun. 224: 431-437, 1996. PubMed ID : 8702406

9. Fajas, L.; Auboeuf, D.; Raspe, E.; Schoonjans, K.; Lefebvre, A. M.; Saladin, R.; Najib, J.; Laville, M.; Fruchart, J.-C.; Deeb, S.; Vidal-Puig, A.; Flier, J.; Briggs, M. R.; Staels, B.; Vidal, H.; Auwerx, J. : The organization, promoter analysis, and expression of the human PPAR-gamma gene. J. Biol. Chem. 272: 18779-18789, 1997. PubMed ID : 9228052

10. Gampe, R. T., Jr.; Montana, V. G.; Lambert, M. H.; Miller, A. B.; Bledsoe, R. K.; Milburn, M. V.; Kliewer, S. A.; Willson, T. M.; Xu, H. E. : Asymmetry in the PPAR-gamma/RXR-alpha crystal structure reveals the molecular basis of heterodimerization among nuclear receptors. Molec. Cell 5: 545-555, 2000. PubMed ID : 10882139

11. Greene, M. E.; Blumberg, B.; McBride, O. W.; Yi, H. F.; Kronquist, K.; Kwan, K.; Hsieh, L.; Greene, G.; Nimer, S. D. : Isolation of the human peroxisome proliferator activated receptor gamma cDNA: expression in hematopoietic cells and chromosomal mapping. Gene Expr. 4: 281-299, 1995. PubMed ID : 7787419

12. Kersten, S.; Mandard, S.; Tan, N. S.; Escher, P.; Metzger, D.; Chambon, P.; Gonzalez, F. J.; Desvergne, B.; Wahli, W. : Characterization of the fasting-induced adipose factor FIAF, a novel peroxisome proliferator-activated receptor target gene. J. Biol. Chem. 275: 28488-28493, 2000. PubMed ID : 10862772

13. Kersten, S.; Desvergne, B.; Wahli, W. : Roles of PPARs in health and disease. Nature 405: 421-424, 2000. PubMed ID : 10839530

14. Kroll, T. G.; Sarraf, P.; Pecciarini, L.; Chen, C.-J.; Mueller, E.; Splegelman, B. M.; Fletcher, J. A. : PAX8-PPAR-gamma-1 fusion in oncogene human thyroid carcinoma. Science 289: 1357-1360, 2000. PubMed ID : 10958784

15. Kubota, N.; Terauchi, Y.; Miki, H.; Tamemoto, H.; Yamauchi, T.; Komeda, K.; Satoh, S.; Nakano, R.; Ishii, C.; Sugiyama, T.; Eto, K.; Tsubamoto, Y.; and 17 others : PPAR-gamma mediates high-fat diet-induced adipocyte hypertrophy and insulin resistance. Molec. Cell 4: 597-609, 1999. PubMed ID : 10549291

16. Lehmann, J. M.; Moore, L. B.; Smith-Oliver, T. A.; Wilkison, W. O.; Willson, T. M.; Kliewer, S. A. : An antidiabetic thiazolidinedione is a high affinity ligand for peroxisome proliferator-activated receptor gamma (PPAR gamma). J. Biol. Chem. 270: 12953-12956, 1995. PubMed ID : 7768881

17. Lowell, B. B. : PPAR-gamma: an essential regulator of adipogenesis and modulator of fat cell function. Cell 99: 239-242, 1999. PubMed ID : 10555139

18. Martin, G.; Schoonjans, K.; Staels, B.; Auwerx, J. : PPAR-gamma activators improve glucose homeostasis by stimulating fatty acid uptake in the adipocytes. Atherosclerosis 137: S75-S80, 1998. PubMed ID : 9694545

19. Meirhaeghe, A.; Fajas, L.; Helbecque, N.; Cottel, D.; Lebel, P.; Dallongeville, J.; Deeb, S.; Auwerx, J.; Amouyel, P. : A genetic polymorphism of the peroxisome proliferator-activated receptor gamma gene influences plasma leptin levels in obese tumors. Hum. Molec. Genet. 7: 435-440, 1998. PubMed ID : 9467001

20. Miles, P. D. G.; Barak, Y.; He, W.; Evans, R. M.; Olefsky, J. M. : Improved insulin-sensitivity in mice heterozygous for PPAR-gamma deficiency. J. Clin. Invest. 105: 287-292, 2000. PubMed ID : 10675354

21. Mueller, E.; Sarraf, P.; Tontonoz, P.; Evans, R. M.; Martin, K. J.; Zhang, M.; Fletcher, C.; Singer, S.; Spiegelman, B. M. : Terminal differentiation of human breast cancer through PPAR-gamma. Molec. Cell. 1: 465-470, 1998. PubMed ID : 9660931

22. Mueller, E.; Smith, M.; Sarraf, P.; Kroll, T.; Aiyer, A.; Kaufman, D. S.; Oh, W.; Demetri, G.; Figg, W. D.; Zhou, X.-P.; Eng, C.; Spiegelman, B. M.; Kantoff, P. W. : Effects of ligand activation of peroxisome proliferator-activated receptor gamma in human prostate cancer. Proc. Nat. Acad. Sci. 97: 10990-10995, 2000. PubMed ID : 10984506

23. Nagy, L.; Tontonoz, P.; Alvarez, J. G. A.; Chen, H.; Evans, R. M. : Oxidized LDL regulates macrophage gene expression through ligand activation of PPAR-gamma. Cell 93: 229-240, 1998. PubMed ID : 9568715

24. Ricote, M.; Huang, J.; Fajas, L.; Li, A.; Welch, J.; Najib, J.; Witztum, J. L.; Auwerx, J.; Palinski, W.; Glass, C. K. : Expression of the peroxisome proliferator-activated receptor gamma (PPAR-gamma) in human atherosclerosis and regulation in macrophages by colony stimulating factors and oxidized low density lipoprotein. Proc. Nat. Acad. Sci. 95: 7614-7619, 1998. PubMed ID : 9636198

25. Ristow, M.; Muller-Wieland, D.; Pfeiffer, A.; Krone, W.; Kahn, C. R. : Obesity associated with a mutation in a genetic regulator of adipocyte differentiation. New Eng. J. Med. 339: 953-959, 1998. PubMed ID : 9753710

26. Rosen, E. D.; Sarraf, P.; Troy, A. E.; Bradwin, G.; Moore, K.; Milstone, D. S.; Spiegelman, B. M.; Mortensen, R. M. : PPAR-gamma is required for the differentiation of adipose tissue in vivo and in vitro. Molec. Cell 4: 611-617, 1999. PubMed ID : 10549292

27. Sarraf, P.; Mueller, E.; Smith, W. M.; Wright, H. M.; Kum, J. B.; Aaltonen, L. A.; de la Chapelle, A.; Spiegelman, B. M.; Eng, C. : Loss-of-function mutations in PPAR-gamma associated with human colon cancer. Molec. Cell 3: 799-804, 1999. PubMed ID : 10394368

28. Tong, Q.; Dalgin, G.; Xu, H.; Ting, C.-N.; Leiden, J. M.; Hotamisligil, G. S. : Function of GATA transcription factors in preadipocyte-adipocyte transition. Science 290: 134-138, 2000. PubMed ID : 11021798

29. Tontonoz, P.; Hu, E.; Devine, J.; Beale, E. G.; Spiegelman, B. M. : PPAR gamma 2 regulates adipose expression of the phosphoenolpyruvate carboxykinase gene. Molec. Cell. Biol. 15: 351-357, 1995. PubMed ID : 7799943

30. Tontonoz, P.; Hu, E.; Graves, R. A.; Budavari, A. I.; Spiegelman, B. M. : mPPAR gamma 2: tissue-specific regulator of an adipocyte enhancer. Genes Dev. 8: 1224-1234, 1994. PubMed ID : 7926726

31. Tontonoz, P.; Hu, E.; Spiegelman, B. M. : Stimulation of adipogenesis in fibroblasts by PPAR-gamma-2, a lipid-activated transcription factor. Cell 79: 1147-1156, 1994. PubMed ID : 8001151

32. Tontonoz, P.; Nagy, L.; Alvarez, J. G. A.; Thomazy, V. A.; Evans, R. M. : PPAR-gamma promotes monocyte/macrophage differentiation and uptake of oxidized LDL. Cell 93: 241-252, 1998. PubMed ID : 9568716

33. Valve, R.; Sivenius, K.; Miettinen, R.; Pihlajamaki, J.; Rissanen, A.; Deeb, S. S.; Auwerx, J.; Uusitupa, M.; Laakso, M. : Two polymorphisms in the peroxisome proliferator-activated receptor-gamma gene are associated with severe overweight among obese women. J. Clin. Endocr. Metab. 84: 3708-3712, 1999. PubMed ID : 10523018

34. Wang, X. L.; Oosterhof, J.; Duarte, N. : Peroxisome proliferator-activated receptor gamma C161-T polymorphism and coronary artery disease. Cardiovasc. Res. 44: 588-594, 1999. PubMed ID : 10690291

35. Yen, C.-J.; Beamer, B. A.; Negri, C.; Silver, K.; Brown, K. A.; Yarnall, D. P.; Burns, D. K.; Roth, J.; Shuldiner, A. R. : Molecular scanning of the human peroxisome proliferator activated receptor gamma (hPPAR-gamma) gene in diabetic Caucasians: identification of a pro12ala PPAR-gamma-2 missense mutation. Biochem. Biophys. Res. Commun. 241: 270-274, 1997. PubMed ID : 9425261